

METRON

RIVISTA INTERNAZIONALE DI STATISTICA — REVUE INTERNATIONALE DE STATISTIQUE
INTERNATIONAL REVIEW OF STATISTICS — INTERNATIONALE STATISTISCHE ZEITSCHRIFT

DIRETTORE PROPRIETARIO — DIRECTEUR ET PROPRIÉTAIRE
EDITOR AND PROPRIETOR — HERAUSGEBER UND EIGENTHÜMER

Prof. Dott. Corrado Gini, *Direttore dell'Istituto di Statistica della R. Università di Roma.*

AMMINISTRATORE — ADMINISTRATEUR — MANAGER — VERWALTER

Prof. Silvio Orlandi, *Istituto di Statistica della R. Università di Roma.*

SEGRETARI DI REDAZIONE — SECRÉTAIRES DE RÉDACTION
EDITORIAL, SECRETARIES — REDACTIONSSECRETAERE

Dott. Ernesto Pizzetti — Dott. Giuseppe Vaccaro

Vol. XIV - N. 1

15-VI-1940.

SOMMARIO — SOMMAIRE — CONTENTS — INHALT

C. Gini. <i>Sur la théorie de la dispersion et sur la vérification des schémas theoriques</i>	Pag. 3
G. Zappa. <i>Osservazioni sopra le medie combinatorie</i>	» 31
E. Pizzetti. <i>Medie ascendenti e medie discendenti</i>	» 55
R. Mogno. <i>Su una formula approssimata per il calcolo di $n!$ e le sue applicazioni</i>	» 67
G. De Meo. <i>Sulle nascite dei primogeniti in Italia</i>	» 79
C. Evelpidi. <i>Le revenu national des pays balkaniques</i>	» 159

ROMA

AMMINISTRAZIONE DEL « METRON »

R. UNIVERSITÀ — ISTITUTO DI STATISTICA

ARTICOLI GIUNTI ALLA RIVISTA CHE
VERRANNO PUBBLICATI NEI PROSSIMI
NUMERI.

(Secondo l'ordine d'arrivo).

ARTICLES REÇUS PAR LA REVUE ET
À PARAÎTRE PROCHAINEMENT.

(D'après la date de reception).

ARTIKEL DIE AN DIE ZEITSCHRIFT
ANGEKOMMEN SIND UND WELCHE IN DEN
NACHFOLGENDEN NUMMERN ERSCHEI-
NEN WERDEN.

(Nach der Reihenfolge des Eingangs).

ARTICLES RECEIVED BY THE REVIEW
WHICH WILL BE PUBLISHED IN FUTU-
RE ISSUES.

(According to date of receipt).

C. Gini. — *Asimmetria e anormalità delle distribuzioni statistiche.*

Gli Autori degli articoli inviati
per la pubblicazione nella Rivista,
rinunciano in favore della medesima
alla proprietà letteraria degli articoli
stessi, qualora vengano pubblicati.

Les Auteurs des articles envoyés à
la Revue, pour y être publiés, renon-
cent, en faveur de celle-ci, à la pro-
priété littéraire de leurs articles, s'ils
sont acceptés.

The Authors of papers sent for
publication in the Review are sup-
posed to give up their copyright in
favour of the Review if the papers
are published.

Die Verfasser der zur Veröffentli-
chung in der Zeitschrift zugesandten
Aufsätze, werden, falls selbige veröf-
fentlicht werden, auf ihre Verfasser-
rechte zu Gunsten der Zeitschrift
verzichten müssen.

CORRADO GINI

Sur la théorie de la dispersion et sur la vérification et l'utilisation des schémas théoriques ⁽¹⁾

Le rapport de fréquence entre naissances masculines et féminines qui, dès le XIV^e ou XIII^e siècle, avait fourni, à Florence, la matière à la constatation d'une loi statistique ⁽²⁾ — peut-être la première loi statistique que l'on ait établie après les tables d'Ulpien sur la vie probable — et auquel une place importante était faite, quelques siècles plus tard, dans les célèbres recherches de John Graunt ⁽³⁾, devait figurer parmi les premières applications aussi du calcul des probabilités et ensuite fournir l'occasion à cette *théorie de la dispersion* qui a marqué une nouvelle orientation dans les dites applications.

Dispersion de certaines grandeurs signifie littéralement leur distribution et en quelques langues le mot est généralement

(1) Rapport présenté à la « Réunion d'études sur l'application du calcul des probabilités » organisée sous le haut patronage du Département Fédéral de l'Intérieur de l'Etat de Genève, par l'Université de Genève et l'Organisation Internationale de Coopération Intellectuelle (Genève, 12-15 Juillet 1939).

(2) Dans la *Cronica* de Jean Villani, du début du XIV^e siècle, on parle de l'excédent des naissances masculines sur les féminines comme d'une régularité que l'on avait déjà établie par une longue observation. « Troviamo dal piovano che battezzava i fanciulli (imperocchè ogni maschio che si battezzava a S. Giovanni, per averne il novero, metteva una fava nera, e per ogni femmina una fava bianca) che erano in questi tempi dalle 55 alle 60 centinaia, avanzando il sesso mascolino che il feminino da 300 in 500 per anno ». Cf. G. VILLANI, *Cronica*, con note di MOUTHIÉ et appendice di DRAGOMANNI, Firenze, 1845.

(3) J. GRAUNT, *Natural and political observations upon the tables of mortality, chiefly with reference to the government, religion, trade growth, air, diseases, etc. of the city of London*, 1661.

employé par les statisticiens comme synonyme de *variabilité*. Mais, dans l'expression *théorie de la dispersion*, on entend par dispersion la comparaison entre la distribution effective des grandeurs considérées et une distribution théorique calculée par le moyen du calcul des probabilités. On parle, de la sorte, de *dispersion normale*, *sous-normale* ou *surnormale*, selon que la distribution effective des grandeurs observées a la même variabilité ou, au contraire, est moins ou plus variable que leur distribution théorique, calculée dans les hypothèses que la probabilité des événements élémentaires (par ex. la probabilité qu'un nouveau-né soit du sexe masculin ou, au contraire, du sexe féminin) reste constante dans tout l'intervalle et dans tout le champ de l'observation, et soit indépendante du fait que ces événements se soient produits ou non dans les cas précédents ⁽¹⁾.

C'est là le schéma attaché au nom de Bernoulli.

A côté de celui-ci d'autres schémas ont été établis, dans lesquels on part d'hypothèses théoriques plus ou moins différentes, soit en tant que les événements élémentaires ne sont pas considérés comme mutuellement indépendants, soit en tant que leurs probabilités ne sont pas considérées comme constantes. Le procédé en tout cas est le même : il consiste à comparer la distribution effective de certaines grandeurs avec une distribution théorique, établie sur la base de certaines hypothèses.

Or, ce procédé a marqué une nouvelle étape dans les applications du calcul des probabilités.

Dans le passé, en effet, le calcul des probabilités, qui était né, ainsi qu'on le sait, pour calculer les combinaisons des jeux de hasard, avait été appliqué, non seulement à ceux-ci, mais aussi aux divers phénomènes physiques et sociaux, en admettant que, pour les dits phénomènes, de même que pour les jeux de hasard,

(1) A la rigueur, il suffirait de dire que la probabilité des événements élémentaires est constante car, si la probabilité d'un événement n'est pas indépendante du fait qu'il se soit produit ou non dans les cas précédents, cette probabilité évidemment varie dans l'intervalle de l'observation. J'ai jugé pourtant de m'en tenir au langage courant — qui, d'ailleurs, est assez clair — d'après lequel on distingue deux ordres de causes qui font varier la probabilité : les causes qui tiennent à des circonstances que l'on pourrait dire extérieures aux événements considérés, et les causes qui dépendent des modalités avec lesquelles les événements eux-mêmes se sont réalisés dans les cas précédents.

on puisse admettre l'hypothèse de la constance des probabilités des événements élémentaires, ainsi que celle de leur indépendance mutuelle. De cette façon, par exemple, était calculée (par les Bernoullis, par Laplace, Poisson, Quételet, Herschel) la probabilité de constater certaines différences entre les rapports de fréquence des deux sexes dans un nombre de naissances donné, ou la probabilité que la fréquence d'un sexe dans un certain nombre de naissances présente un écart de sa probabilité élémentaire supérieur à une certaine limite. Et des applications semblables étaient faites aux erreurs d'observation, à la distribution des coups dans le tir à la cible, à la théorie mécanique de la chaleur et à la théorie cynétique des gas, aux décisions des juges et aux autres résolutions rendues à la pluralité des voix, et ainsi de suite. Dans plusieurs cas, comme dans le cas du rapport des sexes dans les naissances, lesdites hypothèses se sont ensuite avérées conformes à la réalité et les applications du calcul des probabilités se sont montrées, par conséquent, justifiées; dans d'autres, comme dans le cas des décisions rendues à la pluralité des voix, ce fut bien le contraire, et les conclusions non justifiées auxquelles en étaient parvenues des personnes même fort éminentes ont entouré la nouvelle science d'une atmosphère de méfiance qui en a entravé peut-être à un certain moment le développement ultérieur. Mais, en tout cas, le procédé était le même: on partait de la connaissance du mécanisme de certains événements et on en déduisait, par le moyen du calcul des probabilités, la distribution de certaines grandeurs qui en dépendaient. J'appelle *déductives* ces applications du calcul des probabilités.

En contraste avec celles-ci, j'appelle *inductives* les applications du calcul des probabilités qui consistent à comparer la distribution effective de certaines grandeurs avec une distribution théorique, établie par le moyen du calcul des probabilités, en partant de certaines hypothèses ⁽¹⁾.

C'est à Cournot que revient la mérite d'avoir réclamé la comparaison des valeurs prévues par les schémas théoriques éta-

(1) La distinction entre *applications déductives* et *applications inductives* du calcul des probabilités a été faite dans notre article *Contributo alle applicazoni statistiche del calcolo delle probabilità*, « Giornale degli Economisti », Décembre 1907. Cf. aussi *Il sesso dal punto di vista statistico*, Istituto di Statistica della R. Università, Roma, Chap. IV.

blis par le moyen du calcul des probabilités aux données empiriques ⁽¹⁾. Dormoy ⁽²⁾ et Lexis ⁽³⁾ ont donné à cette comparaison un caractère systématique dans la théorie de la dispersion, qui constitue l'exemple classique des applications inductives du calcul des probabilités.

Les deux catégories d'applications — déductives et inductives — ont chacune leur but ; mais il s'agit de buts foncièrement différents. Les applications déductives ont le but de faire des prévisions. Les applications inductives ont un double but : d'une part, un but formel : celui de vérifier le schéma théorique et de permettre par conséquent les applications déductives ; d'autre part, un but substantiel : celui de permettre de pénétrer le mécanisme qui règle le phénomène.

Si, au point de vue historique, les applications inductives sont venues après les applications déductives, au point de vue logique elles doivent, en règle générale, les précéder. Ce n'est qu'exceptionnellement, en effet, que l'on peut faire des applications déductives, avec certitude de succès, sans avoir préalablement vérifié l'exactitude du schéma que l'on veut appliquer. Cela arrive lorsqu'il s'agit d'événements dont nous avons nous-mêmes construit le mécanisme, de façon que nous connaissons à la perfection les conditions dans lesquelles ils se développent. C'est le cas pour les jeux de hasard. Même dans ce cas, d'ailleurs, nous ne pouvons pas nous passer toujours de toute vérification. Quelquefois nous pouvons nous en passer seulement si nous nous contentons d'une prévision qui reste dans un certain degré d'approximation ; mais, si nous aspirons à des prévisions très précises, nous devons procéder à une vérification : c'est le cas, par exemple, dans la *roulette*, dans laquelle bien que tous les numéros aient en théorie la même probabilité de sortir, certains d'entre eux peuvent en réalité sortir avec plus de fréquence à cause de défauts de construction. On

(1) A. COURNOT, *Exposition de la théorie des chances et des probabilités*, Hachette, Paris, 1843.

(2) DORMOY, « Journal des Actuaires Français », 1874, et *Théorie mathématique des assurances sur la vie*, 1878.

(3) W. LEXIS, *Das Geschlechtsverhältniss der Geborenen und die Wahrscheinlichkeitsrechnung*, « Jahrbücher für Nat.-Oek. u. Stat. », 1876 ; *Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft*, Freiburg i. B., 1877 ; *Ueber die Theorie der Stabilität statistischer Reihen*, « Jahrbücher für Nat.-Oek. u. Stat. », 1879.

sait que des joueurs rusés, doués d'un esprit particulier d'observation, ont su dans le passé profiter des « habitudes » des roulettes de Monte-Carlo en réalisant des gains très considérables et en obligeant la direction du Casino à changer, depuis lors, chaque jour les cadrans de la roulette. C'est que notre capacité technique ne sait pas toujours réaliser exactement les conditions requises.

Il est indifférent, au point de vue logique, si on construit d'abord le schéma théorique et on passe ensuite à sa vérification par les faits, ainsi qu'on le fait souvent dans le domaine de la physique et en général des sciences expérimentales ; ou si, au contraire, on part de l'observation d'une certaine distribution effective et on la compare ensuite à un schéma théorique, ainsi qu'on le fait plus souvent dans les sciences sociales et en général dans les sciences d'observation. Pourvu qu'une comparaison entre le schéma théorique et les faits soit opérée, on est toujours dans le domaine des applications inductives du calcul des probabilités.

* * *

Lorsqu'on connaît les probabilités des événements élémentaires — comme dans les applications de la théorie classique de la dispersion — on peut calculer d'une part l'intensité de la variabilité et, d'autre part, la forme de la distribution théorique et comparer l'une et l'autre à la variabilité et à la forme de la distribution effective. Mais la forme de la distribution peut être obtenue aussi sans connaître les probabilités élémentaires. De là la possibilité de comparer aux distributions théoriques, non seulement les distributions effectives qui concernent des *fréquences*, ou, comme on les appelle aussi, des *grandeurs intensives*, mais pareillement celles qui concernent des *montants* ou, comme on les appelle aussi, des *grandeurs extensives*, telles que la taille, le poids, etc. Des applications aux grandeurs extensives ont été même faites par Quételet - avant que la théorie de la dispersion proprement dite eût été formulée. Dans ce cas, nous sommes aussi dans le domaine des applications inductives du calcul des probabilités, car il s'agit de comparaisons entre des distributions effectives et des distributions théoriques déduites par le moyen du calcul des probabilités. Ici aussi, on appelle normales les distributions effectives qui sont conformes aux distributions théoriques établies par le moyen du calcul des probabilités ; mais le mot « normal »

— il faut y faire bien attention — a, dans ce cas, une signification différente que dans la théorie de la dispersion, la conformité à la théorie concernant exclusivement la forme de la distribution et non l'intensité de la variabilité ⁽¹⁾.

* * *

Des schémas théoriques peuvent naturellement être construits non seulement par le moyen du calcul des probabilités, mais aussi en faisant recours à d'autres branches des mathématiques et même sans faire recours aux mathématiques ou, tout au moins, aux mathématiques supérieures, ainsi qu'il arrive souvent dans les schémas construits par les économistes, par les sociologues, par les ethnologistes, etc.

La distinction que nous avons établie entre applications inductives et déductives du calcul des probabilités peut en effet être étendue à tous les schémas théoriques. On parlera de *schémas déductifs* lorsque, croyant bien connaître le mécanisme des événements, on construit un schéma pour prévoir le développement des dits événements, sans se soucier de le vérifier, et en admettant *a priori* qu'à part des circonstances perturbatrices, il doit correspondre à la réalité. On parlera, au contraire, de *schémas inductifs* lorsque ceux-ci sont construits en vue d'être comparés aux faits et ne sont admis que dans la mesure où les faits les vérifient.

(1) Voir, à ce sujet, *Il sesso* etc., Op. cit., pages 84-85. Pour éviter tout danger de confusion, je proposais d'appeler « régulières » les courbes qui ont la *forme* de la distribution correspondant au schéma de Bernoulli et « normales » les courbes qui ont aussi le *degré* de dispersion, c'est-à-dire l'intensité de la variabilité, correspondant au dit schéma. Cette nomenclature pourtant n'a pas été adoptée, et il faut reconnaître qu'en effet l'emploi de l'expression « distribution normale » dans le sens de « distribution gaussienne » est trop répandu pour pouvoir être facilement changé. Gardant par conséquent au mot « normal » cette signification courante, il y aurait lieu de qualifier avec un autre mot les distributions qui présentent une « dispersion normale ». Peut-être serait-il opportun de parler plutôt de « dispersion binomiale » et par conséquent de « dispersion sous-binomiale » ou « sur-binomiale », d'autant plus que les distributions des grandeurs intensives que l'on considère sont bien souvent plus ou moins asymétriques, en contraste avec la symétrie de la courbe gaussienne. Voir dans ce sens, l'article : *Asimmetria e anormalità delle distribuzioni statistiche* qui paraîtra dans le prochain numero de « Metron ».

On fait encore un ample usage des schémas déductifs dans les sciences morales qui se basent sur la psychologie humaine (psychologie, économie politique, sociologie, etc.), car on croit pouvoir se rendre compte avec précision suffisante du mécanisme de la conduite humaine par le moyen de l'introspection. Les résultats de la psychologie et de la sociologie modernes, qui tendent à donner de plus en plus d'importance dans la conduite humaine aux facteurs subconscients et inconscients, sont vraisemblablement destinés à réduire de plus en plus (sinon à éliminer complètement) les schémas déductifs du domaine des sciences morales.

En tout cas, il est bien certain que dès maintenant un rôle prépondérant doit être reconnu, dans toutes les sciences, aux schémas inductifs.

* * *

Pour les schémas inductifs deux groupes de problèmes se posent, correspondant aux deux buts que ces schémas se proposent :

1°. Quand pouvons-nous regarder comme vérifié un schéma théorique ?

2°. Quelles sont les conclusions sur le mécanisme du phénomène que l'on peut tirer d'un schéma inductif vérifié ?

* * *

Combien de fois n'est-il pas arrivé que des schémas théoriques que l'on regardait comme parfaitement vérifiés, et par conséquent certainement correspondant à la réalité, se sont démontrés ensuite inadéquats et ont nécessité des révisions, des corrections ou des intégrations ? Cela dépend souvent du fait que des phénomènes nouveaux se sont produits, ou que d'autres, déjà existants, ont été pris en considération plus tard, auxquels le schéma s'est démontré inadéquat. Il n'y a rien de surprenant ou d'alarmant en cela. C'est la conséquence naturelle du progrès de nos connaissances, auquel nos constructions théoriques doivent nécessairement s'adapter. D'autres fois, c'est qu'au début on se contentait d'une certaine approximation de la théorie aux faits, tandis qu'ensuite on a demandé une approximation supérieure. C'est le cas, par exemple, pour les lois de Newton sur l'attraction universelle.

Là aussi on ne pourrait voir que l'effet de la progression des exigences scientifiques et techniques. D'autres fois encore, c'est que l'on a trouvé que le schéma théorique s'adaptait bien aux cas sur lesquels on l'avait d'abord vérifié, mais ne s'adaptait pas, ou s'adaptait moins bien, à d'autres cas des mêmes phénomènes auxquels on l'a appliqué ensuite et qui, même s'ils étaient apparemment équivalents, devaient contenir des éléments différents. Cette constatation est plus troublante. Comme on ne peut jamais avoir la certitude que deux cas soient parfaitement égaux, on pourrait en effet en déduire que l'on ne peut jamais avoir confiance dans l'application d'un schéma théorique à des cas nouveaux, malgré les vérifications nombreuses et satisfaisantes que l'on en a faites aux mêmes phénomènes. Ce serait là une façon pessimiste de présenter la vérité. Plus exactement, on peut dire que lorsqu'un schéma théorique a été vérifié d'une façon satisfaisante, on peut l'appliquer aux cas nouveaux avec la même confiance avec laquelle on peut regarder les cas nouveaux comme essentiellement équivalents aux cas sur lesquels le schéma a été vérifié dans le passé.

Mais quand pouvons-nous dire qu'un schéma a été vérifié d'une façon satisfaisante ?

On répondra : lorsque les faits correspondent aux données théoriques que le schéma fait prévoir. Il est facile pourtant de montrer que cette condition n'est pas suffisante. Supposons que l'on veuille représenter un phénomène qui présente un certain nombre de valeurs (par exemple, le nombre des mariages dans les sept jours de la semaine), ou bien un phénomène dont nous connaissons un certain nombre de valeurs (par exemple, la population d'un pays pendant un siècle dans les années du recensement qui a lieu tous les dix ans). Supposons encore que, pour représenter ces phénomènes, on ait construit des schémas comportant un nombre de paramètres égal au nombre de valeurs dont nous disposons (dans nos exemples, respectivement 7 ou 10 paramètres). Les données calculées d'après le schéma théorique reproduiront exactement les données observées. On ne pourrait demander une concordance plus précise entre les faits et la théorie. Et, si nous faisons d'autres applications du schéma aux mêmes phénomènes, c'est-à-dire aux mariages des divers jours de la semaine, ou à la population d'un siècle déterminée par des recensements décennaux, nous sommes certains que les données établies d'après le schéma reproduiront les données empiriques.

Malgré cela, nous ne pourrions pas dire d'avoir fait une vérification du schéma. Dans des cas semblables, la comparaison des faits aux données théoriques n'ajoute même rien à notre connaissance.

Pourquoi ?

La raison en est que le schéma est une reproduction simplifiée de la réalité : un schéma qui ait le même degré de complication que la réalité ne mérite pas même la dénomination de schéma et il ne présente aucune utilité pratique. C'est le cas du premier exemple, relatif aux mariages dans les 7 jours de la semaine. Si la réalité est, au contraire, plus compliquée, mais nous en connaissons seulement quelques données dont le nombre est du même ordre de grandeur que le nombre des inconnues qui rentrent dans le schéma, nous pourrions parler d'un schéma, mais nous n'aurons pas les moyens de le vérifier, faute d'éléments suffisants. C'est le cas du second exemple, concernant les recensements décennaux. Dans les deux exemples, nous ne pourrions pas parler de vérification du schéma théorique par les faits.

Pour parler en effet de la « vérification d'un schéma », il faut que les deux cas soient possibles, que le schéma soit vérifié par les faits ou qu'il ne le soit pas. Sans cela, nous pourrions parler d'un schéma « compréhensif », mais non d'un schéma vérifié.

Il y a lieu de se demander si la catégorie des schémas compréhensifs qui ne sont pas vérifiables est limitée aux schémas qui comportent un nombre de paramètres égal au nombre des données qui sont connues, ou si l'on ne serait pas justifié à faire rentrer dans la même catégorie d'autres cas dans lesquels le schéma porte à des formules tellement simples que l'on voit immédiatement que toute série d'événements peut s'y adapter. Comme il y a des affirmations tellement évidentes qui ne font avancer en rien nos connaissances et que nous appelons « platitudes » ou « vérités de M. de La Palisse », de même n'y aurait-il pas des schémas qui ne font avancer en rien nos connaissances, tellement ils sont compréhensifs ? Une remarque de ce genre a été faite au sujet de la logistique généralisée par les Professeurs Pearl et Reed. La logistique originairement établie par Verhulst présente une allure caractéristique qui peut répondre ou peut ne pas répondre aux faits ; mais avec les termes complémentaires et les cycles successifs considérés par Pearl et Reed, la courbe est devenue tellement souple qu'elle s'adapte évidemment à une série de données quelconques lorsqu'on

prend des termes suffisamment éloignés, de façon à éliminer les discontinuités. Il n'est pas surprenant, par conséquent, si, dans de telles conditions, on a trouvé que la dynamique de la population des différents pays est représentée avec de bonnes approximations par la logistique généralisée. Mais il s'en faut que l'on puisse regarder le schéma comme applicable aux développements successifs des populations respectives, même en faisant abstraction de faits nouveaux. Il est facile, au contraire, de démontrer que l'on parviendrait souvent à des résultats absurdes. Prenons, par exemple, le cas de la France et de l'Allemagne dans leurs frontières de 1914. La logistique généralisée décrit bien la dynamique des deux populations ; elle permet aussi de déterminer le maximum de la population, qui — abstraction faite de faits nouveaux — pourrait être atteint par les deux populations : ce maximum était calculé par Pearl à 42 millions et demi pour la France et à 116 millions et demi ou 119 millions et demi pour l'Allemagne ⁽¹⁾. Le résultat était évidemment inadmissible. La surface des deux territoires étant à peu près la même et les ressources du territoire français n'étant pas du tout moindres que celles du territoire allemand, il serait impossible de justifier une telle différence. La différence dépend du fait que, pour adapter la logistique aux faits, il faut placer la période considérée beaucoup plus près de la limite supérieure pour la France que pour l'Allemagne. On arrive à des résultats également incompatibles entre eux pour plusieurs autres pays de l'Europe. Une autre jolie contradiction serait sortie — à ce qu'on m'a dit — de l'application de la logistique aux populations de l'Etat de New York, d'une part, et de la ville de New York, de l'autre. Les données dont on disposait pour les deux populations suivaient admirablement la logistique : le schéma aurait été donc vérifié pour les deux séries ; mais, à partir d'une certaine date future, la population de la ville de New York aurait dépassé celle de l'Etat tout entier et le maximum compatible avec les conditions actuelles aurait donc été supérieur pour la ville que pour l'Etat, dont elle n'est qu'une partie.

On pourrait pourtant donner une autre explication de l'absurdité de certaines prévisions auxquelles conduit la logistique, une

(1) Cf. R. PEARL, *Studies in Human Biology*, William and Wilkins, Baltimore, 1924, pages 603-607 et 609.

explication qui me paraît plus plausible : c'est que les vérifications que l'on fait de ces schémas sont en réalité incomplètes et, de ce fait, seulement apparentes.

En effet, le schéma dont on déduit la courbe logistique porte à déterminer non seulement l'allure de la courbe, mais aussi son origine, d'une part, et, d'autre part, son maximum, qui représente le niveau le plus élevé que la population pourrait atteindre dans les conditions actuelles. Pour vérifier complètement le schéma, il faudrait vérifier aussi cette origine et ce maximum : or, on ne possède pas les données qui permettent de le faire d'une façon exacte. La vérification du schéma est donc incomplète. Dans certains cas, d'ailleurs, comme dans les exemples que nous venons de citer, concernant la France et l'Allemagne, ou l'Etat et la ville de New York, les données que l'on possède ou le rapport de la partie au tout entre les deux populations considérées sont suffisants pour montrer que les niveaux maxima prévus pour les deux pays sur la base de la logistique ne sont pas compatibles entre eux et que par conséquent l'un d'eux, tout au moins, n'est pas admissible.

Il serait exagéré de conclure de ces considérations que la logistique généralisée est dépourvue de toute valeur ; mais on peut en conclure qu'elle n'a pas de valeur en tant que schéma représentatif du développement de la population ; sa valeur est limitée comme formule interpolatrice.

Mais quand pourra-t-on dire que la vérification d'un schéma est complète ?

Absolument complète, elle le serait quand on aurait constaté que toutes les déductions et prévisions que l'on peut tirer du schéma sont vérifiées par les faits avec toute précision et dans toutes circonstances. Cette constatation est en pratique bien difficile, sinon tout à fait impossible. Elle impliquerait, d'ailleurs, que le schéma concerne seulement des phénomènes du passé. Toute application déductive en serait exclue.

Ce qu'on peut demander pratiquement c'est que l'on vérifie, avec un degré de précision suffisant, les déductions et prévisions que l'on peut tirer du schéma sur les faits qui nous sont connus ou tout au moins sur ceux que nous jugeons être en connexion avec le schéma lui-même. C'est ce que l'on n'avait pas fait avec la logistique, en négligeant de vérifier si les maximums qu'elle faisait prévoir pour les différentes populations, dans les conditions actuel-

les, étaient compatibles avec les faits et entre eux. Les faits qui nous sont connus dépendent de l'état de notre patrimoine scientifique ; le degré de précision de la vérification dépend des buts que nous avons en vue dans l'utilisation du schéma ; les connexions entre les faits et le schéma sont en partie évidentes, en partie elles dépendent de la subtilité du chercheur. Ces circonstances expliquent — ainsi que nous l'avons dit — comment des schémas regardés comme vérifiés à un certain moment tombent en défaut à un moment successif, et demandent des intégrations ou doivent même être écartés.

Mais, ces circonstances et toutes autres étant égales, il est évident que, le but pratique de la vérification étant d'autoriser des applications déductives du schéma dans des cas nouveaux, la vérification aura d'autant plus de valeur qu'il y a plus de probabilité que l'accord que l'on a constaté entre le schéma et les faits observés se répète dans des cas nouveaux, et cette probabilité sera d'autant plus élevée que les faits pris en considération pour la vérification sont plus nombreux, variés, complexes en comparaison avec le schéma, ou, en d'autres mots, que le schéma est plus simple en comparaison des faits que l'on a utilisé pour la vérification. Si le schéma comporte un nombre de paramètres égal au nombre des valeurs qui servent à la vérification, l'accord entre schéma et faits est — ainsi que nous l'avons remarqué — inévitable : dans ce cas, cet accord n'autorise qu'une bien faible confiance dans un accord analogue du schéma avec des cas nouveaux. En employant l'expression « degré de liberté » dans un sens un peu large, on pourrait dire que plus grand est le nombre des degrés de liberté, vis-à-vis du schéma, de l'ensemble des faits qui ont servi à la vérification, plus grande est la probabilité que l'accord constaté entre les faits et le schéma ne soit pas accidentel et plus grande est, par conséquent, la probabilité que l'accord se répète sur des faits nouveaux.

A côté du but, important au point de vue pratique, d'autoriser des applications déductives, le schéma a l'utilité, importante au point de vue scientifique, de donner une représentation simplifiée de la réalité. Or il est évident que sa valeur sera d'autant plus élevée que cette simplification est plus grande. Voilà pourquoi les statisticiens plus expérimentés — je serais tenté de dire les véritables statisticiens — restent froids devant des formules établies pour représenter certaines distributions (par exemple, celle des revenus) qui comportent trois, quatre, cinq paramètres. Il

n'est pas surprenant que de telles formules représentent assez bien lesdites distributions ; et, s'il était besoin de faire recours à des interpolations plus précises que celles obtenues par les formules plus simples, bien connues, il serait tout à fait justifié de les employer ⁽¹⁾. Mais, en dehors de cet emploi, on ne voit pas quelle utilité pratique on pourrait tirer des dites formules.

Il y aurait une autre utilité si à chaque paramètre de la formule — ou tout au moins à quelqu'un d'entre eux — on pouvait faire correspondre un certain groupe de circonstances ou un certain aspect de la courbe. Ces différents paramètres représenteraient alors autant d'indices statistiques, dont chacun servirait à donner une mesure d'un certain groupe de circonstances ou d'un certain aspect du phénomène. L'utilité de la formule serait alors de comparer chaque aspect ou chaque groupe de circonstances dans des distributions différentes. La formule de la courbe gaussienne contient deux paramètres, dont un représente la moyenne et l'autre l'écart quadratique moyen. Dans une formule que j'ai proposée pour mesurer la concentration de la prolixité des mariages, il y a aussi deux paramètres dont un représente la prolixité moyenne et l'autre mesure l'inégalité de distribution de ladite prolixité. Dans les formules bien connues, également à deux paramètres, proposées par Pareto pour représenter la distribution des revenus et par moi pour représenter leur concentration, un paramètre (α ou respectivement δ) est fonction de l'inégalité de distribution des revenus et est couramment employé comme indice de cette inégalité ⁽²⁾. Mais cette correspondance devient de plus en plus difficile au fur et à mesure que le nombre des paramètres augmente. Elle n'existe plus dans les formules à trois ou plus paramètres, proposées pour représenter les revenus.

(1) Il est bon d'ajouter que, lorsque les interpolations obtenues par des formules simples ne sont pas suffisamment précises, il est à conseiller, en général, de faire recours à une interpolation graphique. En réalité, les courbes qui comportent de nombreux paramètres ne sont pas même employées dans la pratique pour des buts d'interpolation.

(2) Pour les formules, proposées pour mesurer la concentration de la prolixité des mariages et la concentration des revenus, voir le récent volume *Memorie di metodologia statistica*, Vol. I, *Variabilità e concentrazione* (Giuffrè, Milano, 1939), où sont reproduites, avec quelques additions, nos mémoires concernant ces sujets.

Nous avons parlé de l'utilité des formules dérivées de certains schémas théoriques en vue de l'interpolation et de l'établissement des indices statistiques afin que l'on n'étende pas notre critique au delà de ses limites véritables. Mais il est bon d'ajouter que cette utilité est tout à fait indépendante de la valeur des schémas comme tels. Les schémas théoriques peuvent donner lieu ou non à des formules utiles aux fins d'interpolation ou aptes à fournir des indices statistiques et, par contre, des indices statistiques et des formules interpolatrices d'une utilité indiscutable sont en usage qui ne présupposent aucun schéma théorique.

* * *

Nous passons maintenant à la deuxième question que nous nous étions posée.

Quand un schéma théorique est dûment vérifié, quelles sont les conclusions que l'on en peut tirer sur le mécanisme des phénomènes ?

Supposez que l'on fasse le raisonnement suivant : Dix chambres de 100 mètres cubes chacune donnent (abstraction faite des murs, planchers, etc.) un édifice de 1000 mètres cubes ; si nous constatons qu'un édifice a des dimensions correspondant à 1000 mètres cubes, nous pourrions donc en conclure qu'il contient dix chambres de 100 mètres cubes chacune. Tout le monde en rirait. Et pourtant les déductions que l'on a tirées de la vérification des schémas théoriques ne sont pas plus valables que ce raisonnement. Ce qui est surprenant c'est que des noms parmi les plus beaux de la théorie statistique et des mathématiques s'y trouvent parfois liés.

A tout seigneur tout honneur ! Nous commencerons par Adolphe Quételet. Le schéma théorique de Quételet était celui de la distribution des valeurs observées d'une grandeur dont les observations étaient affectées par des erreurs accidentelles ; ou, d'une façon plus générale, celui de la distribution des grandeurs dépendant d'une cause constante et de multiples causes perturbatrices, variant l'une indépendamment de l'autre. C'a été un des grands mérites de Quételet d'avoir vérifié la validité de ce schéma pour plusieurs caractères anthropométriques de l'homme (taille, poids, force) ; ses résultats ont été, à ce point de vue, confirmés par les chercheurs qui ont suivi. Mais Quételet

ne s'arrêtait pas là ; la conformité des distributions des caractères anthropométriques à la courbe binomiale permet de reconnaître — pensait-il — si les individus appartiennent à un même type et ne sont différenciés que par des causes fortuites. C'est ce qui aurait lieu en ce qui concerne les individus d'une même nation ayant le même âge, les oscillations dans leur taille ou dans les autres caractères étant déterminées — d'après Quételet — par l'influence du climat, la différence des nourritures ou des fatigues plus ou moins grandes ⁽¹⁾. Aujourd'hui même, l'interprétation de Quételet est encore partagée par certains théoriciens : dans un traité bien connu sur « Le hasard », dû à un mathématicien éminent, il est souligné que « le caractère biométrique intrinsèque des séries normales est la pureté de la race ». La série normale des tailles des Français, en particulier, « coïncide — d'après l'auteur — avec le fait qu'ils forment un groupe biologique suffisamment homogène » ⁽²⁾. Tout anthropologiste vous dira, au contraire, que les Français, ainsi d'ailleurs que les populations de beaucoup d'autres pays de l'Europe, sont extrêmement hétérogènes au point de vue racial. Et les biologistes, qui ont démontré expérimentalement, sans contestation possible, que les différences d'une population à distribution binomiale dépendent en partie de facteurs du milieu, tels que ceux envisagés par Quételet, mais en partie aussi de facteurs génétiques, citent volontiers ces conclusions pour démontrer l'insuffisance de la statistique dans l'analyse des phénomènes biologiques ⁽³⁾. En réalité, la faute n'est pas à l'insuffisance de la statistique, mais plutôt à la défaillance de statisticiens ou de mathématiciens éminents (*quandoque bonus dormitat Homerus*), qui ont pris une condition suffisante pour une condition nécessaire. Car le schéma d'une cause constante et de multiples causes accidentelles est théoriquement condition suffisante, mais non condition nécessaire pour la distribution normale, qui peut très bien provenir de mécanismes différents. On peut vérifier cette proposition pour l'espèce humaine en mettant ensemble les mensu-

(1) Cf. *Lettres sur la théorie des probabilités*, Bruxelles, 1846, Lettre XXX, particulièrement, à la page 216, et *Anthropométrie*, Bruxelles, 1871, *Introduction*, particulièrement aux pages 15 et 26.

(2) E. BOREL, *Le hasard*, Paris, 1928, pag. 151.

(3) Cf. W. JOHANNSEN, *Elemente der exakten Erblichkeitslehre*, III, deutsche Auflage, Jena, 1926, particulièrement aux pages 177-186.

rations que l'on possède des tailles, ou des indices céphaliques, de toutes les races. On obtient précisément une courbe qui peut bien être regardée comme binomiale ⁽¹⁾.

Revenons à la théorie de la dispersion. On a vérifié que la dispersion des pourcentages des mâles dans les naissances humaines, groupées d'après les divers intervalles de temps et les divers territoires, est normale, c'est-à-dire que ces pourcentages se comportent *comme si* la probabilité qu'une naissance soit du sexe masculin était constante et indépendante du sexe des naissances précédentes. Ici aussi, n'ont pas manqué ceux qui en ont conclu que la probabilité de la naissance d'un enfant de sexe masculin *est* constante et indépendante du sexe constaté aux naissances précédentes. Lexis lui-même, auteur de la théorie de la dispersion, sans arriver jusque là, s'approchait pourtant beaucoup de cette conclusion, en admettant que la dite probabilité puisse seulement varier d'une femme à l'autre, par des fluctuations purement accidentelles. De la dite dispersion normale des rapports des sexes dans les naissances, il concluait que le sexe de l'enfant est déjà déterminé dans l'ovaire de la femme et le rapport entre ovules de sexe masculin et ovules de sexe féminin est le même chez toutes les femmes ou varie seulement par des fluctuations accidentelles. Les différents rapports entre garçons et filles, que l'on observe dans les divers groupes des naissances — légitimes et illégitimes, nés à la campagne et en ville — dépendraient à son avis du différent pourcentage d'avortements, parmi lesquels — ainsi qu'il est bien connu — le sexe masculin est plus largement représenté ⁽²⁾.

Cette conclusion a été contredite par la génétique, qui a démontré que le sexe des naissances humaines est déterminé à la fécondation et dépend des spermatides; non moins que par la statistique, qui démontre qu'il y a certainement des différences significatives entre les rapports des sexes des diverses familles ⁽³⁾

(1) Voir, à ce sujet, ce que je disais dans *Il sesso* etc., Op. cit., pages 169-170, et dans l'article *Quelques considérations au sujet de la construction des nombres-indices des prix et des questions analogues*, « Metron », Vol. IV, N. 1, 15-VII-1924.

(2) Cf., pour ces interprétations, *Il sesso* etc., Op. cit., pages 149-150.

(3) La part qui, dans les différences entre les rapports des sexes des diverses familles, revient au hasard et la part qui, au contraire, est significative, ont été mesurées dans l'ouvrage cité *Il sesso* etc., Chap. X: *La variabilità individuale nella tendenza a produrre i due sessi*, pages 371-393.

et probablement aussi des divers groupes sociaux. Ici encore on a confondu une condition suffisante avec une condition nécessaire. Il est vrai, en effet, qu'une dispersion normale a lieu lorsque :

1^o la probabilité d'une naissance masculine est constante pour toutes les naissances et indépendante du sexe des naissances précédentes.

Mais il est aussi vrai qu'une dispersion normale a lieu également dans les deux conditions suivantes ;

2^o la probabilité qu'une naissance soit de sexe masculin est indépendante du sexe des naissances précédentes et ne varie pas d'un intervalle à l'autre et d'un territoire à l'autre, mais varie, avec une dispersion normale, d'un groupe à l'autre des naissances du même intervalle ou du même territoire, ces groupes, pourtant, étant, eux aussi, variables en nombre avec une dispersion normale ;

3^o la probabilité qu'une naissance soit de sexe masculin n'est pas indépendante du sexe des naissances précédentes ; mais la tendance, qui en dérive, à une dispersion sous-normale ou sur-normale est compensée par une tendance en sens contraire provenant des variations de la probabilité d'une naissance masculine, soit d'un intervalle ou d'un territoire à l'autre, soit d'un groupe à l'autre des naissances dans le même intervalle ou dans le même territoire (1).

Ayant démontré que la probabilité d'une naissance masculine varie d'une famille à l'autre, la 1^{re} hypothèse reste exclue et ayant, d'autre part, démontré que la probabilité d'une naissance masculine est indépendante du sexe des naissances précédentes (2), la 3^{ème} hypothèse reste également exclue. On en déduit que, dans le cas du sexe des naissances, la dispersion normale correspond à la 2^{ème} hypothèse. De cette conclusion on a pu tirer des conséquences d'une certaine importance, car la constance, dans l'espèce humaine, de la probabilité qu'une naissance soit de sexe masculin d'un intervalle à l'autre et d'un territoire à l'autre porte à écarter toutes les théories sur la détermination du sexe dans notre espèce qui attribuaient de l'importance aux facteurs psychologiques, climatiques, alimentaires, économiques et, en général, aux facteurs du

(1) Cf., à ce sujet aussi, *Il sesso*, etc., Op. cit., particulièrement aux pages 152-156.

(2) Cette démonstration est donnée dans l'ouvrage cité *Il sesso* etc., pages 159-166.

milieu qui sont normalement variables dans le temps et dans l'espace ⁽¹⁾.

Voilà, donc, que des conclusions — même importantes — peuvent être tirées des schémas théoriques vérifiés. On n'y arrive pourtant que par des recherches supplémentaires et cela même dans les cas les plus simples, tels que celui de la théorie de la dispersion.

Que dire de ces schémas compliqués, cumulant hypothèses sur hypothèses, qui — tout au moins en Italie — font, depuis plusieurs années, le plaisir innocent d'un certain groupe de statisticiens, ou de soi-disant statisticiens, et qui devraient — dans leur intention — nous rendre compte, par exemple, de la distribution des revenus ou du développement de la population, phénomènes pour lesquels les schémas plus simples proposés par le passé ne seraient pas satisfaisants ?

Dans le meilleur des cas, on pourra dire que ces schémas, une fois qu'ils fussent dûment vérifiés par les faits, représenteraient une des explications possibles des phénomènes auxquels ils sont appliqués. Je dis « dans le meilleur des cas », car quelquefois les hypothèses qu'ils impliquent sortent du domaine du vraisemblable ⁽²⁾. Mais, même dans le cas le plus favorable, ces schémas — une fois qu'ils fussent dûment vérifiés par les faits — ne pourraient être regardés comme la seule explication admissible qu'après que toutes les hypothèses qu'ils impliquent (ou tout au moins toutes les hypothèses sauf une) eussent été vérifiées séparément.

Si ma pensée n'est pas claire, je la rendrai claire — je l'espère — en reprenant l'exemple des chambres et de l'édifice que j'ai mentionné au début de ma critique.

Nous avons fait le schéma d'un édifice de dix chambres, chacune de 100 mètres cubes et avons ensuite vérifié qu'un certain

(1) Cf. *Il sesso* etc., pages 166-168. Cela n'exclut pas que des circonstances anormales — soit naturelles, soit expérimentales — puissent provoquer des variations significatives dans la probabilité qu'une naissance soit de sexe masculin. C'est ce qui est arrivé, dans beaucoup de pays, pendant et après la guerre, pour des causes qui sont encore incertaines.

(2) Cf., au sujet de quelques schémas proposés pour expliquer la distribution des revenus, ce que je disais dans les *Memorie di metodologia statistica*. Op. cit., pages 110-111. Des observations analogues pourraient être faites au sujet d'autres schémas proposés ensuite.

édifice a précisément les dimensions correspondant aux 1000 mètres cubes que ce schéma comporte. Après cela nous ne pouvons dire autre chose sinon qu'il est possible que cet édifice contienne dix chambres de 100 mètres cubes chacune ; mais, afin de pouvoir dire qu'il les contient vraiment, il faut que nous vérifions ultérieurement : que l'édifice contient 10 chambres ; que 9 des dix chambres ont une capacité de 100 mètres cubes ; la seule vérification dont nous pouvons nous passer — en ayant vérifié les dimensions de l'édifice — est celle de la capacité de la dixième chambre. Alors, et seulement alors, nous serons autorisés à dire que le schéma envisagé correspond à la structure de l'édifice en question.

En général, si un certain schéma comporte n hypothèses totalement ou partiellement indépendantes, et s'il est vérifié, (c'est-à-dire si les résultats qu'il fait prévoir sont confirmés par les faits), on ne pourra jamais dire que le schéma représente la structure des phénomènes considérés sans avoir vérifié aussi $n-1$ des n hypothèses que le schéma comporte ⁽¹⁾.

Bien souvent, pourtant, il ne suffit pas de vérifier $n-1$ hypothèses ; il faut vérifier aussi la $n^{\text{ème}}$ hypothèse.

(1) La confusion entre les conditions à remplir pour la vérification du schéma et celles qui sont nécessaires pour connaître la structure du phénomène a été probablement facilitée par le fait qu'en général un schéma est déclaré représenter une « hypothèse » et, par conséquent, on pense que, lorsque cette hypothèse est vérifiée, il ne reste plus rien à faire. Or, on peut admettre que le schéma représente une hypothèse, pourvu que l'on ait soin de préciser qu'il s'agit d'une « hypothèse complexe », c'est-à-dire d'un « ensemble de plusieurs hypothèses ». Il s'ensuit que, après avoir vérifié l'hypothèse complexe, il reste à vérifier aussi sa composition dans les diverses hypothèses qu'elle comprend. On pourrait même soutenir que le schéma ne peut pas, à la rigueur, être considéré comme vérifié si l'on n'a pas vérifié aussi les diverses hypothèses qu'il comprend : il y a lieu, en effet, de remarquer qu'en acceptant le schéma, on est aussi autorisé à prévoir les faits qui découlent de ces diverses hypothèses, qui, par conséquent, doivent être également soumises à la confirmation des faits. Ce point de vue plus rigoureux a été adopté dans mon discours sur *I Pericoli della Statistica* (1^{ère} réunion de la Société Italienne de Statistique, Pisa, 9 octobre 1939), qui a été écrit après, mais imprimé avant le présent article (Cf. « Rivista di Politica Economica », Roma, Novembre 1939 et « Atti della Riunione di Pisa della Società Italiana di Statistica », Padova 1940).

Dans l'exemple précédent, nous n'avions pas besoin de vérifier la capacité de la 10^{ème} chambre, parce qu'il était possible de la déterminer par différence entre la capacité totale de l'édifice et celle des autres 9 chambres. Mais cela n'arrive pas toujours.

Nous le montrerons par un exemple.

Le prestige de la courbe des erreurs accidentelles est tel qu'il y a eu, et il y a encore, des statisticiens, ou des presque statisticiens, qui attachent une grande valeur à étudier les transformations que l'on doit introduire dans une courbe non normale pour la réduire à une courbe normale. Sauf des cas exceptionnels qui n'ont pas d'intérêt de notre point de vue, il est toujours possible en théorie et il est, en règle générale, possible en pratique, avec des transformations opportunes, de réduire une courbe quelconque à une autre courbe quelconque, de façon que ce but est, en règle générale, atteint avec satisfaction de l'auteur. Jusqu'ici, rien de mal. Mais le mal commence lorsque l'auteur pense avoir construit, avec la dite transformation, un schéma qui donne l'explication de la courbe qui n'est pas normale. Prenons un exemple. La courbe des revenus globaux est du type hyperbolique ; les courbes des qualités individuelles sont du type binomial et par conséquent on peut admettre, en particulier, que la courbe des aptitudes à gagner un revenu soit une courbe du type binomial.

Dans ces conditions, on peut étudier la transformation à apporter dans la courbe des aptitudes à gagner un revenu pour obtenir la courbe des revenus globaux. N'avons-nous pas obtenu de la sorte un schéma qui nous donne l'explication de la courbe des revenus globaux ? La courbe des revenus globaux — pourra-t-on observer — déduite du schéma correspond aux faits ; le schéma est donc vérifié. Des deux hypothèses que le schéma comporte — distribution binomiale des aptitudes à gagner un revenu, relation entre cette aptitude et le revenu global représentée par la transformation — une, la première, est vérifiée. Ne pouvons-nous pas en conclure qu'aussi la deuxième est vérifiée, comme nous avons déduit la capacité de la 10^{ème} chambre de la capacité globale de l'édifice et de la capacité des autres chambres ? Pas du tout ! car il y a une seule capacité de la 10^{ème} chambre qui puisse accorder la capacité globale connue de tout l'édifice avec les capacités connues des 9 chambres ; mais il y a, au contraire, plusieurs relations qui peuvent concilier la distribution du type hyperbolique des revenus globaux avec la distribution du type

binomial des aptitudes à gagner un revenu. Pour expliquer la distribution des revenus globaux, il faut donc vérifier aussi la deuxième hypothèse, c'est-à-dire la relation entre aptitude à gagner un revenu et revenu global, relation qui correspond à la transformation opérée. Or, la dite transformation correspond à l'hypothèse qu'entre aptitude et revenu global il y ait une relation biunivoque, ce qui n'est pas du tout conforme à la vérité. A égalité d'aptitudes à gagner un revenu, le revenu global, en effet, change d'après l'âge, les relations et maintes autres circonstances, en premier lieu la fortune. Entre phénomènes statistiques les relations biunivoques sont l'exception ; peut-être n'en existe-t-il même pas ⁽¹⁾. C'est là la raison pour laquelle les transformations des courbes ne peuvent pas, dans le domaine statistique, servir à établir des schémas théoriques répondant à la réalité.

Il n'est pas du tout impossible d'établir la relation entre aptitudes à gagner un revenu et revenu global ; il suffit de construire une table à double entrée pour les deux phénomènes. De cette façon on pourra parvenir à un schéma qui explique la courbe des revenus globaux au moyen de la courbe des aptitudes à gagner un revenu.

Revenons encore une fois à la dispersion du rapport des sexes dans les naissances.

Ici le schéma était vérifié : la dispersion était en effet normale. Le schéma comportait deux hypothèses : *a*) probabilité du sexe masculin constante dans toutes les naissances ; *b*) indépendance entre le sexe d'une naissance et le sexe des naissances précédentes. La deuxième hypothèse a été aussi vérifiée. Mais il aurait été erroné d'en déduire que, par conséquent, la première hypothèse aurait aussi pu être regardée comme vérifiée. En effet, la dispersion normale peut également être conciliée avec l'hypothèse *b*) par une hypothèse différente de l'hypothèse *a*) ; c'est-à-dire par l'hypo-

(1) Dans la statistique on a pourtant l'occasion d'utiliser parfois des relations biunivoques entre phénomènes qui ne sont pas statistiques, telle, par exemple, la relation qui passe entre les dioptries et les distances focales des lentilles. La dioptrie est l'inverse de cette distance. On peut, par une transformation, passer de la distribution d'un groupe d'individus selon les dioptries des lunettes qu'ils emploient (ou devraient employer) à leur distribution selon la distance de l'oeil du point éloigné (*punctum remotum*) réel ou virtuel. Contre une transformation pareille il n'y a rien à objecter.

thèse *c*) d'une probabilité variable dans des groupes d'extension variable. Il devenait, par conséquent, nécessaire de vérifier aussi l'hypothèse *a*). La vérification a été négative. De là on a pu conclure à la réalité de l'hypothèse *c*) et arriver de la sorte à des connaissances positives sur le mécanisme des facteurs du rapport des sexes dans les naissances humaines.

Cette analyse de la dispersion du rapport des sexes dans les naissances remonte à plus de 30 ans.

Une analyse semblable devrait être faite, ou tout au moins tentée, toutes les fois que l'on vise à utiliser les schémas théoriques pour accroître nos connaissances sur le mécanisme des phénomènes. Malheureusement, on ne le fait que trop rarement.

On continue, au contraire, à fabriquer des schémas parfois assez compliqués qui se cumulent, s'enchevêtrent, parfois se contredisent, sans prendre soin de vérifier sur les faits les hypothèses qu'ils comportent ou même d'examiner leur plausibilité.

On dit que l'on a satisfait de la sorte au besoin humain de se donner une explication des phénomènes. Il n'y a pas de doute que c'est là un mobile très noble et très utile ; c'est le mobile qui est à la base de la recherche scientifique ; mais, pourqu'il porte à des résultats scientifiques, il faut qu'il soit dirigé et réglementé par une procédure rigoureuse. La vérité est que construire des schémas est très facile et peut donner l'illusion d'avoir fait des découvertes ; examiner et vérifier les hypothèses qu'ils comportent est long et pénible, et bien souvent cela a le résultat de faire disparaître les illusions que l'on s'était faites. On a dit des recherches scientifiques en général, mais on peut le répéter à plus forte raison pour les schémas théoriques, qu'ils sont comme les enfants, dont l'engendrement est rapide et agréable, alors que la gestation est longue et laborieuse. Qui voudrait pousser l'analogie pourrait comparer les faiseurs de schémas théoriques qui se désintéressent des vérifications nécessaires à les rendre viables, à ces pères qui, après avoir contribué à l'engendrement de l'enfant, se soustraient à la responsabilité qu'ils ont encourue. Pour ma part, je me bornerai à les comparer à des expérimentateurs impatientes et paresseux qui entreprennent beaucoup de recherches, mais n'en portent aucune à terme. Peut-être satisfont-ils de la sorte leur curiosité, mais ils n'enrichissent pas notre patrimoine scientifique. Et encore moins aboutissent-ils à simplifier et éclaircir la réalité ; loin de là, ils ne font que répandre l'incertitude et créer la confusion. Cela

au point de vue de la théorie. Les inconvénients ne sont pas moindres au point de vue pratique, car les déductions que l'on pourrait faire de ces schémas théoriques, croyant qu'ils nous renseignent vraiment sur le mécanisme du phénomène, peuvent nous porter loin de la réalité. Il est vrai que tôt ou tard les faits se chargeront de nous faire rentrer en nous-mêmes. Mais il n'est pas nécessaire d'attendre de donner de la tête contre le mur pour se convaincre d'avoir fait fausse route. Un petit écart au point de départ peut porter à des écarts énormes au point d'arrivée, si on ne corrige pas les déviations en cours de voyage. C'est pour cela que tout bon navigateur fait périodiquement le point. C'est cela aussi que doit faire tout bon chercheur, ne se fiant pas aux constructions théoriques, mais faisant appel aux faits pour les contrôler dans la mesure du possible. Moins de schémas théoriques et plus de vérifications étendues, détaillées et approfondies des hypothèses : voilà un programme qui s'impose dans certains secteurs des sciences sociales.

Il est vrai que bien des fois on n'a pas les éléments pour vérifier toutes les hypothèses que les différents schémas alternatifs comportent, surtout s'il s'agit de schémas compliqués, de façon que l'on reste incertain sur leur validité et on doit baser son choix sur des raisons extra-scientifiques, telles que la simplicité du schéma, son utilité pratique, son harmonie avec notre conception générale de l'univers ou tout simplement avec nos sentiments personnels. Mais il ne faut pas ajouter, aux incertitudes inévitables, des incertitudes que l'on pourrait éviter avec un peu plus de discipline mentale et un peu plus de travail matériel.

La tâche de la vérification des diverses hypothèses que le schéma comporte est — je le disais — d'autant plus difficile que le schéma est plus compliqué. Cette considération s'ajoute à celles que nous avons faites dans les pages précédentes pour plaider en faveur des schémas simples.

* * *

Une autre observation à faire concerne la multiplicité des schémas théoriques établis pour les mêmes phénomènes. Pas tous les schémas différents sont incompatibles ; au contraire, pour les phénomènes statistiques, qui dépendent en général d'un nombre très grand d'autres phénomènes, la possibilité de construire plusieurs schémas différents, mais compatibles entre eux, est la règle.

Nous avons vu comment on peut construire un schéma qui rende compte de la distribution des revenus globaux en fonction de la distribution des aptitudes à gagner ; analogiquement, on pourrait construire un schéma qui rende compte de la distribution des revenus globaux en fonction de la distribution des fortunes. Il va de soi qu'on pourrait établir un schéma plus compliqué en fonction, en même temps, des aptitudes à gagner et des fortunes ; mais on pourrait de même établir un autre schéma qui mette en rapport la distribution des revenus avec la distribution par âge, par profession, par instruction, par sexe.

Tout schéma théorique, ainsi que nous l'avons remarqué, est une représentation simplifiée de la réalité. Or, les simplifications peuvent être faites de différentes façons ; et différentes simplifications peuvent être également autorisées. Mais alors — peut-on observer — même après avoir vérifié le schéma dans ses résultats globaux et avoir vérifié également les différentes hypothèses (toutes ou toutes moins une) qu'il comporte, on ne peut pas dire qu'on soit arrivé à l'explication du phénomène ; il y aura toujours d'autres explications également autorisées. Oui et non - c'est la réponse. On peut dire qu'on est arrivé à l'explication du phénomène en fonction de certains autres phénomènes ; mais il y aura la possibilité d'autres explications en fonction de phénomènes partiellement ou totalement différents.

* * *

Il y a d'autres dangers dans la construction des schémas théoriques : c'est d'oublier une ou quelques-unes des hypothèses qu'ils comportent. C'est un danger auquel on est plus exposé lorsqu'il s'agit de schémas compliqués, mais dont les schémas simples ne sont pas tout à fait exempts. Quoi de plus simple que le schéma envisagé plus haut, celui d'un édifice de 1000 mètres cubes composé de 10 chambres de 100 mètres cubes chacune ? et pourtant, dans son énoncé et dans son application, nous avons oublié l'hypothèse qu'il fût bâti de plein pied, de façon que sa capacité correspondît à ses dimensions extérieures. Au cas où cette hypothèse ne serait pas réalisée, la conclusion sur la capacité de la dixième chambre, à laquelle nous sommes arrivés, et par conséquent la connaissance de la structure de l'édifice, que nous avons tirée de l'application du schéma, seraient erronées.

Dans les cas envisagés dans les paragraphes précédents, les conclusions erronées, tirées des schémas théoriques, provenaient du fait que nous prenions des conditions suffisantes pour des conditions nécessaires ; ici, l'erreur proviendrait du fait que nous aurions négligé une des conditions nécessaires pour arriver à nos conclusions.

C'est là peut-être une des causes les plus fréquentes qui exigent la révision d'une théorie que nous avons construite pour expliquer les phénomènes et qui avait semblé pour un certain temps tout à fait satisfaisante ; c'est que la théorie contenait des hypothèses qui étaient passées inaperçues. Pour autant que la vérification avait été faite dans des conditions ou relativement à des cas qui étaient conformes à ces hypothèses inaperçues, la théorie fonctionnait d'une façon satisfaisante ; lorsqu'on est passé à l'application dans des conditions ou à des cas différents, ces défauts se sont rendus manifestes. Ces diverses conditions ou ces cas différents, auxquels on a étendu la vérification, peuvent bien rentrer dans ces faits nouveaux auxquels est due, ainsi qu'il est reconnu et que nous l'avons mentionné d'abord, la modification de nos schémas théoriques. Bien souvent les schémas qui se sont de la sorte révélés insuffisants ne doivent pas être rejetés ; ils peuvent être conservés en leur apportant les compléments nécessaires.

Les dites insuffisances sont fréquentes surtout lorsqu'on part de l'observation d'une certaine régularité et que l'on construit ensuite le schéma destiné à l'expliquer. Il arrive facilement, dans ce cas, de construire un schéma qui explique seulement certaines caractéristiques de la régularité observée, sans rendre compte des autres. C'est ce qui est arrivé dans les premiers schémas de la courbe des revenus. Sur la base des recherches de Galton, mettant en lumière l'uniformité de la distribution des aptitudes individuelles, on avait prévu l'uniformité de la distribution des revenus. Pareto, ayant démontré la différence essentielle entre les deux distributions, l'expliquait par les difficultés que l'on éprouve à passer d'une classe sociale à l'autre. Ces difficultés étant insuffisantes à donner une raison de la forte asymétrie de la courbe des revenus, Benini faisait entrer en ligne de compte l'institution de la succession, qui a sans doute une certaine importance, mais ne suffit pas non plus à expliquer l'allure caractéristique de la courbe. Nous avons alors signalé une autre circonstance, qui peut nous en rendre compte, dans la plus forte reproduction des classes sociales plus pauvres.

Il est même douteux si, en ces cas, on peut proprement parler de schémas théoriques ou s'il ne vaudrait pas mieux de les appeler tout simplement des explications.

La différence entre schéma et explication est, en effet, principalement une différence de degré. Tout schéma n'est, au fond, qu'une explication complète de la structure du phénomène. Il a pourtant une autre caractéristique : c'est d'être une explication qui concerne la physionomie générale du phénomène, et non pas les particularités qualitatives et quantitatives avec lesquelles il se manifeste dans des cas particuliers. Du fait d'être des explications générales et complètes proviennent les avantages des schémas théoriques : ils expliquent la faveur croissante que ceux-ci rencontrent dans les milieux scientifiques modernes.

* * *

L'examen critique que nous venons de faire des schémas théoriques conduit à des conclusions qui ne sont pas excessivement encourageantes. Tout à fait exceptionnellement, les schémas théoriques peuvent être employés sans vérification préalable : c'est le cas seulement lorsque nous avons construit nous-mêmes le mécanisme des phénomènes considérés, comme dans les jeux de hasard, et quelquefois, dans ce cas même, seulement si on se contente d'une certaine approximation. Dans tous les autres cas, les schémas théoriques doivent être vérifiés au moyen d'une comparaison adéquate des résultats qu'ils font prévoir avec la réalité. Mais pas toujours les schémas sont vérifiables, soit parce qu'ils ne présentent aucune simplification vis-à-vis de la réalité (et alors ils ne méritent pas même le nom de schémas), soit parce que les données pour une vérification adéquate manquent. D'ailleurs, bien des fois la vérification n'est qu'incomplète, et, de ce fait, à un certain point de vue, illusoire. Dans tous ces cas, le schéma ne peut servir qu'à des buts éventuels d'interpolation entre les données connues. Lorsqu'une vérification adéquate du schéma est faite, le schéma ne présente qu'un ensemble d'hypothèses qui doivent être vérifiées séparément — toutes ou, dans quelques cas, toutes moins une — pour nous autoriser à tirer des conclusions sur le véritable mécanisme du phénomène. Lorsque toutes ces précautions sont dûment observées, nous pouvons arriver à nous rendre compte du mécanisme du phénomène relativement à

certaines circonstances, le schéma étant foncièrement une simplification de la réalité qui doit négliger d'autres circonstances. Ces précautions ne sont en réalité observées que très exceptionnellement et l'emploi des schémas théoriques a, par conséquent, donné lieu — même de la part de statisticiens et mathématiciens des plus éminents — à des conclusions erronées qui n'ont pas contribué à élever le prestige de la statistique. D'ailleurs, lorsque les schémas sont compliqués, il serait bien difficile, sinon impossible, de procéder à toutes les vérifications qui seraient nécessaires ; de façon que cet instrument de recherche, ainsi qu'il est employé aujourd'hui, se montre souvent plus dangereux qu'utile.

La faute n'est pas aux schémas théoriques ; mais à la façon, trop simpliste, de s'en servir. Les schémas théoriques ne sont que des explications de la structure des phénomènes, mais des explications qui devraient être complètes et générales. Ils ont marqué certainement un véritable progrès dans la méthodologie scientifique. Mais il s'agit d'instruments délicats ; on ne doit pas les employer à la légère. Il faut faire moins de schémas théoriques, mais se soumettre à un travail sérieux de vérification, soit du schéma lui-même dans son ensemble, soit des diverses hypothèses qu'il comporte. Il faut faire des schémas aussi simples que possible ; non seulement le schéma simple remplit mieux sa fonction, qui est de donner une représentation simplifiée des phénomènes, mais il est plus facile d'en faire une vérification complète, ou moins incomplète ; cette vérification, les circonstances étant égales, a plus de validité, et la vérification également des hypothèses que le schéma comporte est plus facile à être réalisée.

Vérifiés d'une façon adéquate et complétés par les recherches supplémentaires nécessaires sur la conformité aux faits, ou tout au moins sur la plausibilité des hypothèses qu'ils renferment, les schémas théoriques peuvent porter — et ont porté en effet — à des conclusions importantes.

G. ZAPPA

Osservazioni sopra le medie combinatorie.

Ci proponiamo di studiare le medie chiamate dal GINI « Medie combinatorie », e precisamente di esaminare anzitutto il comportamento di tali medie nel caso di serie con ripetizione ; di definire di conseguenza delle medie combinatorie ponderate ; di definire inoltre le medie combinatorie con ripetizione ; di dare alcune interessanti proprietà delle medie combinatorie con e senza ripetizione.

Inoltre, a simiglianza di quanto è stato già fatto da vari autori per le medie potenziate, cerchiamo di estendere la definizione di media combinatoria al caso di una distribuzione continua della variabile ; e infine, vogliamo determinare il limite a cui tende una media combinatoria di una serie i cui termini siano in progressione aritmetica, geometrica, quadratica, armonica, etc., quando, fermi restando i termini estremi, il numero dei termini tende all'infinito.

I. CONCETTO DI COMBINAZIONE MEDIA D'ORDINE c , DI UNA SERIE DI n TERMINI ($n \geq c$).

Sia data una serie di n termini positivi, a_1, a_2, \dots, a_n . Non facciamo, per ora, nessuna ipotesi sopra la possibilità o meno che alcuni termini siano eguali tra loro.

Chiameremo, per comodità di linguaggio, *combinazione semplice d'ordine c* della serie ($c \leq n$) il prodotto di c termini distinti della serie stessa.

Diremo *combinazione media semplice d'ordine c di 1ª specie* della serie stessa la media aritmetica delle $\binom{n}{c}$ combinazioni semplici d'ordine c della serie.

Diremo invece *combinazione media semplice d'ordine c di 2^a specie* la media aritmetica di quelle tra le $\binom{n}{c}$ combinazioni semplici d'ordine c che hanno i c termini non solo distinti, ma anche con valore diverso.

In tal caso c dovrà non superare il numero dei termini differenti della serie.

Es.: Sia data la serie di 5 termini

$$2, 2, 3, 5, 6$$

Le $\binom{5}{2} = 10$ combinazioni semplici d'ordine 2 della serie sono le seguenti

$$\begin{array}{ccccccccc} 2 \times 2 = 4 & 2 \times 3 = 6 & 2 \times 5 = 10 & 2 \times 6 = 12 & 2 \times 3 = 6 & 2 \times 5 = 10 & & & \\ & 2 \times 6 = 12 & 3 \times 5 = 15 & 3 \times 6 = 18 & 5 \times 6 = 30 & & & & \end{array}$$

La combinazione media semplice d'ordine 2 di 1^a specie è

$$\frac{4 + 6 + 10 + 12 + 6 + 10 + 12 + 15 + 18 + 30}{10} = \frac{123}{10} = 12,3$$

Per ottenere la combinazione media semplice d'ordine 2 di 2^a specie della serie non bisogna invece tener conto della combinazione 2×2 , risultando essa di due termini di eguale valore. Tale combinazione media sarà data da

$$\frac{6 + 10 + 12 + 6 + 10 + 12 + 15 + 18 + 30}{9} = \frac{119}{9} = 13,2$$

Chiameremo inoltre *combinazione con ripetizione d'ordine c della serie*, il prodotto di c termini distinti o coincidenti della serie stessa. Non è naturalmente, più necessario supporre $c \leq n$.

Diremo *combinazione media con ripetizione d'ordine c di 1^a specie* la media aritmetica delle $\binom{n+c-1}{c}$ combinazioni con ripetizione d'ordine c della serie.

Diremo infine *combinazione media con ripetizione d'ordine c di 2^a specie* la media aritmetica di quelle tra le $\binom{n+c-1}{c}$ combinazioni con ripetizione d'ordine c della serie, in cui tra i c termini non ve ne siano due formalmente distinti, ma aventi valore eguale.

Es.: Riprendendo la serie di 5 termini

2, 2, 3, 5, 6

le $\binom{6}{2} = 15$ combinazioni con ripetizione sono date, oltre che dalle 10 combinazioni semplici, dalle seguenti

$$2 \times 2 = 4 \quad 2 \times 2 = 4 \quad 3 \times 3 = 9 \quad 5 \times 5 = 25 \quad 6 \times 6 = 36$$

La combinazione media con ripetizione d'ordine 2 di 1^a specie è pertanto

$$\frac{4 + 6 + 10 + 12 + 6 + 10 + 12 + 15}{15} + \frac{18 + 30 + 4 + 4 + 9 + 25 + 36}{15} = \frac{201}{15} = 13,4$$

Per ottenere la combinazione media con ripetizione d'ordine 2 di 2^a specie non bisogna tener conto della combinazione ottenuta moltiplicando il primo 2 per il secondo 2. Tale combinazione media pertanto sarà:

$$\frac{6 + 10 + 12 + 6 + 10 + 12 + 15 + 18}{14} + \frac{30 + 4 + 4 + 9 + 25 + 36}{14} = \frac{197}{14} = 14,1$$

2. DEFINIZIONE DEI VARI TIPI DI MEDIE COMBINATORIE:

Partendo dal concetto di combinazione media, ora introdotto, potremo definire in modo molto semplice le medie combinatorie, introdotte, almeno nelle loro forme più generali, dal GINI (1).

Diremo *media combinatoria monoplana semplice d'ordine c di 1^a specie* la radice c-esima della combinazione media semplice d'ordine c di prima specie. Essa coincide con la media chiamata dal Gini « media combinatoria monoplana d'ordine c ».

Diremo *media combinatoria monoplana semplice d'ordine c di 2^a specie* la radice c-esima della combinazione media semplice d'ordine c di 2^a specie.

Diremo così *media combinatoria monoplana con ripetizione d'ordine c di 1^a (2^a) specie* la radice c-esima della combinazione media con ripetizione d'ordine c di 1^a (2^a) specie.

(1) C. GINI, *Di una formula comprensiva delle medie*, « Metron », vol. XIII, n. 1, pagg. 3-22.

Scriviamo ora le espressioni analitiche di queste medie. Sia data una serie, con p_1 termini eguali ad a_1 , p_2 termini eguali ad a_2, \dots, p_n termini eguali ad a_n .

La media combinatoria monoplana semplice d'ordine c di 1^a specie della suddetta serie è data dalla formula

$$M_1^c = \sqrt[c]{\frac{\sum_s \binom{p_1}{q_1^{(s)}} \binom{p_2}{q_2^{(s)}} \dots \binom{p_n}{q_n^{(s)}} a_1^{q_1^{(s)}} a_2^{q_2^{(s)}} \dots a_n^{q_n^{(s)}}}{\sum_s \binom{p_1}{q_1^{(s)}} \binom{p_2}{q_2^{(s)}} \dots \binom{p_n}{q_n^{(s)}}}}$$

dove il sistema $q_1^{(s)}, q_2^{(s)}, \dots, q_n^{(s)}$, descrive, al variare di s , tutti i possibili sistemi di numeri interi tali che

$$0 \leq q_i^{(s)} \leq p_i \quad (i = 1, 2, \dots, n) \quad \sum_{i=1}^n q_i^{(s)} = c$$

La media combinatoria monoplana semplice d'ordine c di 2^a specie M_2^c della suddetta serie è data dalla formula medesima di quella di prima specie, con la differenza che in tal caso il sistema $q_1^{(s)}, q_2^{(s)}, \dots, q_n^{(s)}$ descrive al variare di s , tutti i possibili sistemi di numeri interi tali che, per ciascun valore di i da 1 ad n , si ha $q_i^{(s)} = 0$ oppure $q_i^{(s)} = 1$, ed inoltre $\sum_{i=1}^n q_i^{(s)} = c$.

La media combinatoria monoplana con ripetizione d'ordine c di 1^a specie della suddetta serie è data dalla formula

$${}_R M_1^c = \sqrt[c]{\frac{\sum_s \binom{p_1 + q_1^{(s)} - 1}{q_1^{(s)}} \binom{p_2 + q_2^{(s)} - 1}{q_2^{(s)}} \dots \binom{p_n + q_n^{(s)} - 1}{q_n^{(s)}} a_1^{q_1^{(s)}} a_2^{q_2^{(s)}} \dots a_n^{q_n^{(s)}}}{\sum_s \binom{p_1 + q_1^{(s)} - 1}{q_1^{(s)}} \binom{p_2 + q_2^{(s)} - 1}{q_2^{(s)}} \dots \binom{p_n + q_n^{(s)} - 1}{q_n^{(s)}}}}$$

ove il sistema di valori $q_1^{(s)}, q_2^{(s)}, \dots, q_n^{(s)}$ descrive, al variare di s , tutti i possibili sistemi di numeri interi positivi o nulli tali che $\sum_{i=1}^n q_i^{(s)} = c$.

La media combinatoria monoplana con ripetizione d'ordine c di 2^a specie della medesima serie è data dalla formula

$${}_R M_2^c = \sqrt[c]{\frac{\sum_s \binom{p_1}{\bar{q}_1^{(s)}} \binom{p_2}{\bar{q}_2^{(s)}} \dots \binom{p_n}{\bar{q}_n^{(s)}} a_1^{q_1^{(s)}} a_2^{q_2^{(s)}} \dots a_n^{q_n^{(s)}}}{\sum_s \binom{p_1}{\bar{q}_1^{(s)}} \binom{p_2}{\bar{q}_2^{(s)}} \dots \binom{p_n}{\bar{q}_n^{(s)}}}}$$

ove il sistema $q_1^{(s)}, q_2^{(s)}, \dots, q_n^{(s)}$ descrive, al variare di s , tutti i possibili sistemi di valori positivi o nulli tali che $\sum_{i=1}^n q_i^{(s)} = c$, ed ove è $\bar{q}_i^{(s)} = 0$ se $q_i^{(s)} = 0$, mentre $\bar{q}_i^{(s)} = 1$ se $q_i^{(s)} \neq 0$.

I valori p_1, p_2, \dots, p_n possono anche interpretarsi, anzichè come frequenze, più generalmente come pesi; in tal caso potremo chiamare la M_1^c , *media combinatoria ponderata di prima specie*, e la M_2^c *media combinatoria ponderata di 2ª specie*, e analogamente, la ${}_R M_1^c$ *media combinatoria ponderata di 1ª specie*, e la ${}_R M_2^c$, *media combinatoria ponderata con ripetizione di 2ª specie*. Dette medie conservano significato pure qualora i pesi p_1, p_2, \dots, p_n non siano interi, purchè però restino positivi e, nella M_1^c , maggiori dell'unità.

Chiameremo *media combinatoria monoplana potenziata semplice (con ripetizione) d'ordini c, p di 1ª (2ª) specie* la radice (c/p) -esima, della combinazione media semplice (con ripetizione) d'ordine c di 1ª (2ª) specie delle potenze p -esime dei termini.

Chiameremo *media combinatoria biplana semplice (con ripetizione) d'ordini c, d di 1ª (2ª) specie* la radice d'ordine $c-d$ del rapporto tra la combinazione media semplice (con ripetizione) d'ordine c di 1ª (2ª) specie e la corrispondente d'ordine d , dei termini della serie.

Diremo *media combinatoria equipotenziata biplana d'ordini c, d, p semplice (con ripetizione) di 1ª (2ª) specie* la radice d'ordine $p(c-d)$ del rapporto tra la combinazione media semplice (con ripetizione) d'ordine c di 1ª (2ª) specie e la corrispondente d'ordine d , delle potenze p -esime dei termini della serie.

Diremo ancora *media biplana equicombinatoria potenziata d'ordini c, p, q semplice (con ripetizione) di 1ª (2ª) specie* la radice d'ordine $c(p-q)$ del rapporto tra la combinazione media semplice (con ripetizione) d'ordine c di 1ª (2ª) specie delle potenze p -esime e quella delle potenze q -esime dei termini della serie.

Diremo infine *media biplana combinatoria potenziata d'ordini c, d, p, q , semplice (con ripetizione) di 1ª (2ª) specie* la radice d'ordine $c(p-dq)$ del rapporto tra la combinazione media semplice (con ripetizione) d'ordine c di 1ª (2ª) specie delle potenze p -esime dei termini della serie e la corrispondente d'ordine d delle potenze q -esime dei termini della serie.

Occorre appena avvertire che le medie semplici di 1ª specie di ciascun tipo corrispondono alle medie considerate dal prof. Gini

nell'opera citata; riteniamo inoltre superfluo scrivere le formule relative ai vari tipi di medie.

3. SU ALCUNE PROPRIETÀ DELLE MEDIE COMBINATORIE.

Le medie combinatorie di prima specie hanno un significato più immediato di quelle di seconda specie. Le medie combinatorie di seconda specie godono però della seguente importante proprietà:

Se si moltiplicano tutti i pesi di una media combinatoria di 2^a specie per uno stesso numero reale positivo il valore della media non cambia.

Basterà, per dimostrare tale proprietà, dimostrare che se si moltiplicano tutti i termini di una combinazione media di 2^a specie per uno stesso numero reale positivo, il valore della combinazione media non cambia.

Riferendoci infatti, p. es. alla combinazione media semplice di 2^a specie d'ordine c , data dalla formula

$$\frac{\sum_s \binom{p_1}{q_1^{(s)}} \binom{p_2}{q_2^{(s)}} \dots \binom{p_n}{q_n^{(s)}} a_1^{q_1^{(s)}} a_2^{q_2^{(s)}} \dots a_n^{q_n^{(s)}}}{\sum_s \binom{p_1}{q_1^{(s)}} \binom{p_2}{q_2^{(s)}} \dots \binom{p_n}{q_n^{(s)}}} \quad (I)$$

(ove il sistema $q_1^{(s)}, q_2^{(s)}, \dots, q_n^{(s)}$ descrive, al variare di s , tutti i possibili sistemi di valori tali che $q_i^{(s)} = 0$ oppure $q_i^{(s)} = 1$

($i = 1, 2, \dots, n$) ed inoltre $\sum_{i=1}^n q_i^{(s)} = c$), si nota subito che,

se si moltiplicano i valori p_1, p_2, \dots, p_n dei pesi per uno stesso numero reale positivo k , i prodotti del tipo $a_1^{q_1^{(s)}} a_2^{q_2^{(s)}} \dots a_n^{q_n^{(s)}}$

restano invariati, mentre le espressioni $\binom{p_i}{q_i^{(s)}}$ restano invariate

($= 1$) se $q_i^{(s)} = 0$, e vengono moltiplicate per k se $q_i^{(s)} = 1$.

Pertanto ciascun termine della sommatoria che compare al numeratore della (I) viene moltiplicato per k^c , e altrettanto avviene di ciascun termine della sommatoria che compare al denominatore. La combinazione media resta così invariata.

Nel caso che avessimo a che fare con una combinazione media con ripetizione di 2^a specie, la dimostrazione si farebbe esattamente allo stesso modo.

Tale proprietà non è valida invece per le medie combinatorie di 1^a specie, come può vedersi per mezzo di semplicissimi esempi.

Sia data, p. es., la serie a_1, a_2 , con pesi rispettivamente 2 e 1. La media combinatoria monoplana semplice di 1^a specie d'ordine 2 di detta serie è data dal valore

$$M_1^2 = \sqrt{\frac{a_1^2 + 2 a_1 a_2}{3}}$$

Se diamo invece ai termini della medesima serie i pesi rispettivamente 4 e 2, proporzionali ai precedenti, si ottiene

$$M_1^2 = \sqrt{\frac{6 a_1^2 + 8 a_1 a_2 + a_2^2}{15}}$$

valore evidentemente diverso dal precedente.

Altra proprietà interessante è la seguente :

La combinazione media con ripetizione d'ordine c di una serie di 1^a (2^a) specie è maggiore o eguale della corrispondente combinazione media semplice della stessa specie, l'uguaglianza valendo solo quando tutti i termini della serie sono eguali tra loro.

Faremo la dimostrazione per le combinazioni medie di 1^a specie; basterebbero poche modifiche per avere la dimostrazione anche per le combinazioni medie di 2^a specie.

Ci riferiremo inoltre, per semplicità alla combinazione media d'ordine 5 di una serie di sette termini, $a_1, a_2, a_3, a_4, a_5, a_6, a_7$. Si noti però che il procedimento dimostrativo ha portata del tutto generale.

Due combinazioni con ripetizione si dicono *simili*, quando si possono ottenere l'una dall'altra permutando gli indici dei termini in modo opportuno, od anche, se si vuole, quando nelle due combinazioni gli esponenti dei termini coincidono a meno dell'ordine. Saranno, per esempio, simili le combinazioni $a_1^3 a_2^2, a_1^2 a_3^2, a_2^3 a_3^2, a_2^2 a_3^2$, etc., o le altre $a_1^3 a_2 a_3, a_1^3 a_4 a_5, a_1 a_2^3 a_3, a_1 a_3^3 a_5$, etc.

Si potranno dividere le combinazioni con ripetizione in *tipi* di combinazioni simili. Le combinazioni appartenenti ad un tipo possono individuarsi dalla successione degli esponenti dei loro termini, disposti in ordine non decrescente. Se, per es., si tratta di combinazioni medie d'ordine 5, i vari tipi di combinazioni simili sono individuati dalle seguenti successioni di esponenti :

$$(5), (4,1), (3,2), (3,1,1), (2,2,1), (2,1,1,1), (1,1,1,1,1)$$

Le combinazioni con ripetizione del tipo (5) sono $a_1^5, a_2^5, a_3^{(5)} \dots$ etc., quelle del tipo (4,1) $a_1^4 a_2, a_1 a_2^4, a_1^4 a_3, a_1 a_3^4$, etc.,

e così via, e infine le combinazioni del tipo $(1, 1, 1, 1, 1)$ sono le combinazioni semplici, vale a dire $a_1 a_2 a_3 a_4 a_5$, $a_1 a_2 a_3 a_4 a_6$, $a_1 a_2 a_3 a_5 a_6$, etc.

Possiamo ordinare i vari tipi nel seguente modo :

Se due tipi hanno differente esponente massimo, precederà tra i due quello in cui l'esponente massimo è maggiore ; se due tipi hanno eguale l'esponente massimo, ma diverso l'esponente immediatamente inferiore al massimo, precederà tra i due quello in cui detto esponente è maggiore, e così via.

Così, tra i tipi $(4, 1)$ e $(3, 2)$ precederà il tipo $(4, 1)$, perchè l'esponente massimo, nel primo è 4, nel secondo è 3 ; e tra i tipi $(3, 2)$ e $(3, 1, 1)$ precederà il tipo $(3, 2)$, perchè in esso, pur essendo eguale ($= 3$) l'esponente massimo, si ha che l'esponente immediatamente inferiore è nel primo 2, nel secondo 1.

Ciò posto, è facile vedere che, sempre nel caso di combinazioni d'ordine 5, l'ordine dei tipi è il seguente

- a) (5) , b) $(4, 1)$, c) $(3, 2)$, d) $(3, 1, 1)$, e) $(2, 2, 1)$,
f) $(2, 1, 1, 1)$, g) $(1, 1, 1, 1, 1)$.

Or bene, vale la seguente proprietà :

« Se un tipo x precede un tipo y , la media delle combinazioni del tipo x è maggiore o eguale di quella delle combinazioni del tipo y , l'uguaglianza valendo solo se tutti i termini sono eguali »

Basterà dimostrare ciò per due tipi consecutivi. Prendiamo, ad es., i tipi $(3, 2)$, e $(3, 1, 1)$. Si noti anzitutto che, considerando eguali a 0 gli esponenti mancanti al numero di 5 nei vari tipi, si può affermare che si passa da un tipo al successivo diminuendo di 1 un esponente, ed aumentandone di 1 un altro ad esso seguente. Così, nel caso dei tipi $(3, 2)$ e $(3, 1, 1)$, si diminuisce di 1 il secondo esponente, e si aumenta di 1 il terzo.

Ciò posto, sia una data combinazione del tipo $(3, 1, 1)$, per es. $a_1^3 a_2 a_3$. Consideriamo, insieme ad essa, le altre combinazioni dello stesso tipo che possono formarsi con le lettere a_1, a_2, a_3 . Esse sono $a_1 a_2^3 a_3$ e $a_1 a_2 a_3^3$.

Consideriamo poi le combinazioni del tipo $(3, 2)$ che si possono fare con le lettere a_1, a_2, a_3 . Esse sono

$$a_1^3 a_2^2, a_1^2 a_2^3, a_1^3 a_3^2, a_1^2 a_3^3, a_2^3 a_3^2, a_2^2 a_3^3.$$

È facile vedere che la somma

$$a_1^3 a_2^2 + a_1^2 a_2^3 + a_1^3 a_3^2 + a_1^2 a_3^3 + a_2^3 a_3^2 + a_2^2 a_3^3$$

è maggiore della somma

$$2 a_1^3 a_2 a_3 + 2 a_1 a_2^3 a_3 + 2 a_1 a_2 a_3^3$$

Si ha infatti, supponendo, ad es., $a_1 \leq a_2 \leq a_3$

$$\begin{aligned} & a_1^3 a_2^2 + a_1^2 a_2^3 + a_1^3 a_3^2 + a_1^2 a_3^3 + a_3^3 a_2^2 + a_2^2 a_3^3 - \\ & - (2 a_1^3 a_2 a_3 + 2 a_1 a_2^3 a_3 + 2 a_1 a_2 a_3^3) = (a_1^3 a_2^2 - a_1^3 a_2 a_3) + \\ & + (a_1^2 a_3^3 - a_1 a_2^3 a_3) + (a_1^3 a_3^2 - a_1^3 a_2 a_3) + (a_1^2 a_3^3 - a_1 a_2 a_3^3) + \\ & + (a_2^3 a_3^2 - a_1 a_2^3 a_3) + (a_2^2 a_3^3 - a_1 a_2 a_3^3) = a_1^3 a_2 (a_2 - a_3) + \\ & + a_1 a_2^3 (a_1 - a_3) + a_1^3 a_3 (a_3 - a_2) + a_1 a_3^3 (a_1 - a_2) + \\ & + a_2^3 a_3 (a_3 - a_1) + a_2 a_3^3 (a_2 - a_1) = (a_2 - a_1) (a_2 a_3^3 - a_1 a_3^3) + \\ & + (a_3 - a_1) (a_2^3 a_3 - a_2^3 a_1) + (a_3 - a_2) (a_1^3 a_3 - a_1^3 a_1) = \\ & = a_3^3 (a_2 - a_1)^2 + a_2^3 (a_3 - a_1)^2 + a_1^3 (a_3 - a_2)^2. \end{aligned}$$

quantità manifestamente ≥ 0 , il segno = valendo solo quando è $a_1 = a_2 = a_3$.

È facile poi da ciò dedurre che la media delle combinazioni del tipo (3, 2) è maggiore o eguale alla media delle combinazioni del tipo (3, 1, 1), l'uguaglianza valendo solo nel caso che tutti i termini siano tra loro eguali. Lasciamo la prova al lettore.

La dimostrazione fatta è di carattere affatto generale; essa è quindi valevole per due qualunque tipi consecutivi, come anche nel caso che non si tratti di combinazioni d'ordine 5.

Da ciò segue, sempre riferendoci al nostro caso, che, essendo il tipo (1, 1, 1, 1, 1) l'ultimo nell'ordinamento stabilito, la media delle combinazioni del tipo (1, 1, 1, 1, 1) è minore o uguale alla media delle combinazioni di un qualunque altro tipo, l'uguaglianza valendo solo quando tutti i termini sono uguali tra loro.

Notando poi che la combinazione media semplice è la media delle combinazioni del tipo (1, 1, 1, 1, 1), mentre la combinazione media con ripetizione è la media delle combinazioni di tutti i tipi, si ha immediatamente, da quanto precede, che la combinazione media semplice è sempre minore o eguale della corrispondente combinazione media con ripetizione, l'uguaglianza valendo solo nel caso in cui tutti i termini siano eguali tra loro, *c.d.d.*

Dalla proprietà ora dimostrata segue subito l'altra:

Ogni media combinatoria potenziata monoplana M^p con ripetizione è maggiore della corrispondente media senza ripetizione se è $p > 0$, è minore di essa se è $p < 0$.

Lo studio delle medie biplane da questo punto di vista è più complesso, ed è pertanto rinviato ad altra occasione.

4. MEDIE COMBINATORIE DI UNA VARIABILE CONTINUA.

Sia una variabile x distribuita con continuità nell'intervallo (a, b) (ove è $a > 0, b > 0$). La funzione $f(x)$, che ci dà le frequenze per i vari valori di x , sia continua e sempre positiva in tutto l'intervallo (a, b) . Quando considereremo medie combinatorie ponderate di 1^a specie, supporremo inoltre che sia sempre $f(x) \geq 1$ per ogni x di (a, b) .

Ci proponiamo di definire le varie medie combinatorie per la suddetta variabile x di frequenza $f(x)$. Dovremo anzitutto definire per le variabili continue il concetto di combinazione media. Procederemo a tal uopo nel seguente modo.

Dividiamo l'intervallo (a, b) in parti, mediante una successione di punti di ascisse a_1, a_2, \dots, a_{k-1} , ove è $a < a_1 < a_2 < \dots < a_{k-2} < a_{k-1} < b$.

Le k parti in cui risulta diviso (a, b) come anche le loro lunghezze si indichino rispettivamente con le lettere $\delta_1, \delta_2, \dots, \delta_k$. Sia δ_r una parte generica, e sia x_r un valore di x tale che $a_{r-1} < x_r < a_r$, tale cioè che il punto di ascissa x_r appartenga all'intervallo δ_r .

Consideriamo la combinazione media d'ordine c di 1^a (2^a) specie con o senza ripetizione della serie discreta costituita dai valori x_r ($r = 1, 2, \dots, k$) con pesi rispettivamente $f(x_r) \delta_r$. Evidentemente, la serie discreta così ottenuta ci darà tanto meglio un'idea della serie continua di partenza, quanto minori saranno le parti δ_r . Detta pertanto \bar{K}^c la combinazione media, del genere considerato, della serie discreta, saremo portati a definire la *combinazione media della serie continua*, K^c , come il limite a cui tende la combinazione media della serie discreta quando tendono a zero tutte le parti δ_r ($r = 1, 2, \dots, k$); in simboli

$$K^c = \lim_{\delta_r \rightarrow 0} \bar{K}^c.$$

Perchè detta definizione sia legittima occorre però che il limite suddetto sia indipendente dalla maniera in cui si è diviso l'intervallo (a, b) , dai valori x_r scelti in ciascun δ_r , e dal modo in cui si fanno tendere a 0 i δ_r stessi.

Ci accingiamo ora a dimostrare ciò, e nello stesso tempo a calcolare detto limite.

Nel caso di una combinazione media d'ordine c di 2^a specie senza ripetizione, si ha

$$(2) \quad \bar{K}^c = \frac{\sum_{i_1=1}^k \sum_{i_2=1}^{i_1-1} \sum_{i_3=1}^{i_2-1} \dots \sum_{i_c=1}^{i_{c-1}-1} \delta_{i_1} \delta_{i_2} \dots \delta_{i_c} f(x_{i_1}) f(x_{i_2}) \dots f(x_{i_c}) x_{i_1} x_{i_2} \dots x_{i_c}}{\sum_{i_1=1}^k \sum_{i_2=1}^{i_1-1} \sum_{i_3=1}^{i_2-1} \dots \sum_{i_c=1}^{i_{c-1}-1} \delta_{i_1} \delta_{i_2} \dots \delta_{i_c} f(x_{i_1}) f(x_{i_2}) \dots f(x_{i_c})}$$

Ora si ha, per definizione di integrale

$$\begin{aligned} l i m_{\delta_r \rightarrow 0} \sum_{i_1=1}^k \sum_{i_2=1}^{i_1-1} \sum_{i_3=1}^{i_2-1} \dots \sum_{i_c=1}^{i_{c-1}-1} \delta_{i_1} \delta_{i_2} \dots \delta_{i_c} f(x_{i_1}) f(x_{i_2}) \dots f(x_{i_c}) x_{i_1} x_{i_2} \dots x_{i_c} = \\ = \int_a^b x_1 f(x_1) dx_1 \int_a^{x_1} x_2 f(x_2) dx_2 \dots \int_a^{x_{c-1}} x_c f(x_c) dx_c \end{aligned}$$

e analogamente

$$\begin{aligned} l i m_{\delta_r \rightarrow 0} \sum_{i_1=1}^k \sum_{i_2=1}^{i_1-1} \sum_{i_3=1}^{i_2-1} \dots \sum_{i_c=1}^{i_{c-1}-1} \delta_{i_1} \delta_{i_2} \dots \delta_{i_c} f(x_{i_1}) f(x_{i_2}) \dots f(x_{i_c}) = \\ = \int_a^b f(x_1) dx_1 \int_a^{x_1} f(x_2) dx_2 \dots \int_a^{x_{c-1}} f(x_c) dx_c \end{aligned}$$

Ora, si vede facilmente che⁽¹⁾

$$\int_a^b x_1 f(x_1) dx_1 \int_a^{x_1} x_2 f(x_2) dx_2 \dots \int_a^{x_{c-1}} x_c f(x_c) dx_c = \frac{1}{c!} \left\{ \int_a^b x f(x) dx \right\}^c$$

e analogamente

$$\int_a^b f(x_1) dx_1 \int_a^{x_1} f(x_2) dx_2 \dots \int_a^{x_{c-1}} f(x_c) dx_c = \frac{1}{c!} \left\{ \int_a^b f(x) dx \right\}^c$$

(1) Ciò discende dalla proprietà nota che se $\varphi(x)$ è una funzione integrabile, si ha

$$\int_a^b \varphi(x_1) dx_1 \int_a^{x_1} \varphi(x_2) dx_2 \dots \int_a^{x_{c-1}} \varphi(x_c) dx_c = \frac{1}{c!} \left[\int_a^b \varphi(x) dx \right]^c \quad (\alpha)$$

Essa diviene banale per $c = 1$; la si può pertanto dimostrare per induzione, ammettendola vera per $c - 1$, e dimostrandola per c . Si ha infatti, che sotto tale ipotesi, il primo membro della (α) diviene

$$\frac{1}{(c-1)!} \int_a^b \varphi(x_1) dx_1 \left[\int_a^{x_1} f(x_2) dx_2 \right]^{c-1}$$

Integrando ora per parti, prendendo $\left[\int_a^{x_1} f(x_2) dx_2 \right]^{c-1}$ per fattore finito e $\varphi(x_1) dx_1$ come fattore differenziale, si arriva alla (α) .

Si ha pertanto

$$K = \frac{\frac{1}{c!} \left\{ \int_a^b x f(x) dx \right\}^c}{\frac{1}{c!} \left\{ \int_a^b f(x) dx \right\}^c} = \left\{ \frac{\int_a^b x f(x) dx}{\int_a^b f(x) dx} \right\}^c$$

Nel caso di combinazioni medie di 2^a specie con ripetizione, come pure nel caso di combinazioni medie di 1^a specie con o senza ripetizione, al numeratore della formula (2) che ci dà \bar{K}^c , come anche al denominatore, si deve aggiungere un altro termine, relativo a tutte le combinazioni in cui figura due volte uno stesso valore x_i . Ma è facile vedere che detto termine è composto di sommatorie relative a $c - 1$ indici al massimo, mentre ciascun addendo delle sommatorie ha c fattori δ_r , distinti o coincidenti, che tendono a zero, ed è quindi infinitesimo d'ordine c quando tendono a zero tutti i δ_r . Da ciò si deduce facilmente che il limite di questi termini aggiunti, quando tutti i δ_r tendono a zero, è zero. Ne segue che, anche nel caso di combinazioni medie di 2^a specie con ripetizione, e in quello di combinazioni medie di 1^a specie con o senza ripetizione, si ha

$$K^c = \left\{ \frac{\int_a^b x f(x) dx}{\int_a^b f(x) dx} \right\}^c$$

Ricordiamo che l'espressione

$$\frac{\int_a^b x f(x) dx}{\int_a^b f(x) dx}$$

ci dà la media aritmetica della variabile x , di peso $f(x)$, in (a, b) , abbiamo che

La combinazione media d'ordine c di 1^a o 2^a specie, con o senza ripetizione, di una variabile x di peso $f(x)$ in (a, b) è la potenza c -esima della sua media aritmetica.

Da questa seguono le altre importanti proprietà:

Le medie combinatorie monoplane di qualunque ordine, (di 1^a o 2^a specie, con o senza ripetizione), di una variabile x distribuita con frequenza $f(x)$ nell'intervallo (a, b) , coincidono con la sua media aritmetica.

Le medie monoplane combinatorie potenziate, di ordini c , p , di una variabile x distribuita con frequenza $f(x)$ nell'intervallo a, b coincidono tutte con la sua media monoplana potenziata d'ordine p .

Le medie combinatorie biplane d'ordini c, d di una variabile x di frequenza $f(x)$ nell'intervallo a, b coincidono tutte con la sua media aritmetica; le medie biplane combinatorie equipotenziate di ordini c, d, p coincidono con la media potenziata monoplana d'ordine p della x ; le medie biplane equicombinatorie potenziate d'ordini c, p, q coincidono con la media biplane potenziata d'ordini p, q della x ; e infine le medie biplane combinatorie potenziate di ordini c, d, p, q coincidono anch'esse con la media biplane potenziata d'ordini p, q della x .

5. LIMITI CUI TENDONO ALCUNE MEDIE COMBINATORIE.

G. GALVANI ⁽¹⁾, ha determinato il limite a cui tende una media monoplana potenziata di una serie di termini x_1, x_2, \dots, x_n , in progressione aritmetica, quando, restando fissi gli estremi $x_1 = a, x_n = b$, il numero dei termini tende all'infinito.

Egli giunge ai seguenti risultati:

Se p è l'ordine della media potenziata considerata, il limite di cui sopra è dato da

$$\left[\frac{b^{p+1} - a^{p+1}}{(p+1)(b-a)} \right]^{\frac{1}{p}} \quad \text{per } p \neq 0 \text{ e } p \neq -1$$

$$\frac{b-a}{\log b - \log a} \quad p = -1$$

$$\frac{1}{e} \left[\frac{b^b}{a^a} \right]^{\frac{1}{b-a}} \quad \text{per } p = 0$$

Vogliamo determinare il medesimo limite per le varie medie combinatorie, ed esaminiamo anzitutto, come al solito, il comportamento delle combinazioni medie.

(1) *Dei limiti a cui tendono alcune medie*, « Boll. Unione Matematica Italiana », 1927.

Siano i valori x_1, x_2, \dots, x_n in progressione aritmetica, e sia $x_1 = a$, $x_n = b$. Dividiamo l'intervallo (a, b) dell'asse x in n parti eguali mediante i punti

$$a_r = a + \frac{b - a}{n} \cdot r \quad (r = 0, 1, 2, \dots, n)$$

Essendo poi

$$x_r = a + \frac{b - a}{n - 1} (r - 1) \quad (r = 1, 2, \dots, n)$$

si vede immediatamente che il punto di ascissa x_r è nell'intervallo (a_{r-1}, a_r) ($r = 1, 2, \dots, n$).

Fissata una combinazione media (di ordine c qualsiasi, di 1^a o di 2^a specie, con o senza ripetizione), dei termini x_1, x_2, \dots, x_n in progressione aritmetica, ci proponiamo di determinare il limite cui essa tende, quando, fissi restando $x_1 = a$ e $x_n = b$, il numero n tende all'infinito.

È facile vedere, in base alla definizione di combinazione media nel continuo, che tale limite coincide con la combinazione media della variabile x distribuita nell'intervallo (a, b) con frequenza 1. Basta infatti, ricordare che la combinazione media nel continuo si ottiene in tal caso dividendo (a, b) in parti δ_r , scegliendo in ciascuna parte un valore \bar{x}_r di x , facendo la combinazione media dei valori \bar{x}_r con pesi δ_r , e passando al limite quando tutti i δ_r tendono a zero. Scegliendo in particolare $\delta_r = (a_{r-1}, a_r)$ ($r = 1, 2, \dots, n$) e $\bar{x}_r = x_r$, si ha la coincidenza della combinazione media continua col limite da noi cercato.

Ma la variabile continua x distribuita con frequenza 1 nell'intervallo (a, b) ha per combinazione media d'ordine c (di 1^a o di 2^a specie, con o senza ripetizione), il valore

$$\left[\frac{\int_a^b x \, d x}{\int_a^b d x} \right]^c = \left(\frac{b + a}{2} \right)^c$$

Avremo pertanto, in base alle considerazioni precedenti, che *il limite cui tende la combinazione media d'ordine c (di 1^a o 2^a specie con o senza ripetizione) della serie x_1, x_2, \dots, x_n i cui termini sono in*

progressione aritmetica, quando, fermi restando $x_1 = a$, $x_n = b$, il numero n dei termini tende all'infinito, è dato dal valore

$$\left(\frac{b+a}{2} \right)^c$$

Da ciò si ricava subito che il limite cui tende una qualsiasi media combinatoria mono-plana o bi-plana è dato da

$$\frac{b+a}{2}$$

cioè dalla semisomma dei termini estremi.

Lo studio delle medie combinatorie potenziate richiede invece qualche ulteriore osservazione.

Lievemente modificando il ragionamento che ci ha condotto, nel § 4, a dimostrare che, in base alla definizione data di combinazione media di una variabile x di frequenza $f(x)$ in (a, b) , si ottiene per detta combinazione media il valore

$$\left[\frac{\int_a^b x f(x) dx}{\int_a^b f(x) dx} \right]^c$$

è facile provare che, volendo analogamente definire per una variabile continua la media potenziata d'ordine $p(\neq 0)$ delle combinazioni d'ordine c dei termini, essa è data dal valore

$$\left[\frac{\int_a^b x^p f(x) dx}{\int_a^b f(x) dx} \right]^{\frac{c}{p}}$$

Ragionando come nel caso della media combinatoria, si vede quindi che la media potenziata d'ordine $p(\neq 0)$ delle combinazioni d'ordine c di n termini x_1, x_2, \dots, x_n in progressione aritmetica tende, quando, fissi restando i termini estremi $x_1 = a$, $x_n = b$, il numero n dei termini tende all'infinito, al valore

$$\left[\frac{\int_a^b x^p dx}{\int_a^b dx} \right]^{\frac{c}{p}}$$

che è eguale al valore

$$\left[\frac{b^{p+1} - a^{p+1}}{(p+1)(b-a)} \right]^{\frac{c}{p}} \text{ per } p \neq -1$$

e al valore

$$\frac{b-a}{\log b - \log a} \text{ per } p = -1$$

Notando infine che la media monoplana combinatoria potenziata d'ordini c , p ($p \neq 0$) (di 1^a o 2^a specie, con o senza ripetizione) è la radice c -esima della media potenziata d'ordine p delle combinazioni d'ordine c dei termini, si ottiene subito che *la media monoplana combinatoria potenziata d'ordini c , p ($p \neq 0$) (di 1^a o 2^a specie, con o senza ripetizione) di una serie di termini x_1, x_2, \dots, x_n in progressione aritmetica tende, quando, restando fissi gli estremi $x_1 = a$, $x_n = b$, il numero n dei termini tende all'infinito, al valore*

$$\left[\frac{b^{p+1} - a^{p+1}}{(p+1)(b-a)} \right]^{\frac{1}{p}} \text{ per } p \neq -1$$

e al valore

$$\frac{b-a}{\log b - \log a} \text{ per } p = -1.$$

Se è poi $p = 0$, la media monoplana combinatoria potenziata si riduce alla media geometrica (1). In tal caso il limite da noi cercato è stato già determinato da GALVANI nella memoria citata, ed è dato da

$$\frac{1}{c} \left[\frac{b^b}{a^a} \right]^{\frac{1}{b-a}}$$

Passando alle medie biplane, si ottiene analogamente che *per le medie biplane equicombinatorie potenziate, di ordini c , p , q (p e $q \neq 0$) tale limite è dato da*

$$\left[\frac{(b^{p+1} - a^{p+1})(q+1)}{(b^{q+1} - a^{q+1})(p+1)} \right]^{\frac{1}{p-q}} \text{ per } p \text{ e } q \neq -1$$

(1) Vedi C. GINI, op. cit.

da

$$\left[\frac{(b^{p+1} - a^{p+1})}{(p+1)(\log b - \log a)} \right]^{\frac{1}{p+1}} \text{ per } p \neq -1, q = -1$$

e da

$$\left[\frac{(q+1)(\log b - \log a)}{b^{q+1} - a^{q+1}} \right]^{\frac{1}{-1-q}} \text{ per } p = -1, q \neq -1$$

Così si ha analogamente che per le medie biplane combinatorie equipotenziate di ordini c, d, p ($p \neq 0$) il limite cercato è dato da

$$\left[\frac{b^{p+1} - a^{p+1}}{(p+1)(b-a)} \right]^{\frac{1}{p}} \text{ per } p \neq -1$$

e da

$$\frac{b-a}{\log b - \log a} \text{ per } p \neq -1$$

Infine, per le medie biplane combinatorie potenziate di ordini c, d, p, q (p e $q \neq 0$) tale limite è dato da

$$\left[\frac{\left\{ \frac{b^{p+1} - a^{p+1}}{(p+1)(b-a)} \right\}^c}{\left\{ \frac{b^{q+1} - a^{q+1}}{(q+1)(b-a)} \right\}^d} \right]^{\frac{1}{c p - q d}} \text{ per } p \text{ e } q \neq -1$$

da

$$\left[\frac{\left\{ \frac{b-a}{\log b - \log a} \right\}^c}{\left\{ \frac{b^{q+1} - a^{q+1}}{(q+1)(b-a)} \right\}^d} \right]^{\frac{1}{-c - q d}} \text{ per } p = -1, q \neq -1$$

e da

$$\left[\frac{\left\{ \frac{b^{p+1} - a^{p+1}}{(p+1)(b-a)} \right\}^c}{\left\{ \frac{b-a}{\log b - \log a} \right\}^d} \right]^{\frac{1}{c p + d}} \text{ per } p \neq -1, q = -1$$

R. CISBANI ⁽¹⁾ ha determinato il limite a cui tende una media monoplana potenziata di una serie di termini in progressioni ar-

(1) *Di alcune medie continue - Contributi alla teoria delle medie*, « *Metron* », Vol. XIII, n. 2, 28 febbraio 1928.

monica, geometrica, quadratica, cubica, etc., quando, restando fissi gli estremi $x_1 = a$, $x_n = b$, il numero dei termini tende all'infinito.

Egli giunge ai seguenti risultati :

Se i termini della serie sono tali che le loro potenze j -esime ($j \neq 0$) sono in progressione aritmetica, il limite cercato, per la media monoplana potenziata d'ordine p , ($p \neq 0$), è dato da

$$\left[\frac{b^{p+j} - a^{p+j}}{\left(\frac{p}{j} + 1\right) (b^j - a^j)} \right]^{\frac{1}{p}} \quad \text{se è } p \neq -j$$

e da

$$\left[\frac{b^j - a^j}{j (\log b - \log a)} \right]^{\frac{1}{j}} \quad \text{se è } p \neq -j$$

mentre per la media geometrica (corrispondente al caso $p = 0$) esso è dato da

$$\frac{1}{e^j} \left[\frac{b^{b^j}}{a^{a^j}} \right]^{\frac{1}{b^j - a^j}}$$

Se invece i termini della serie sono in progressione geometrica (corrispondentemente al caso $j = 0$) il limite cercato è dato da

$$\left[\frac{b^p - a^p}{p (\log b - \log a)} \right]^{\frac{1}{p}}$$

per la media monoplana potenziata d'ordine p ($p \neq 0$), e da

$$\sqrt{a b}$$

per la media geometrica (corrispondente al caso $p = 0$).

Naturalmente, per $j = 1$, si ricade nei risultati di GALVANI.

È interessante esaminare la questione, anzichè per le medie potenziate, per le medie combinatorie.

Tenendo presente il ragionamento da noi fatto nel caso di termini in progressione aritmetica, e quello fatto dal CISBANI nella memoria citata, si giunge facilmente ai seguenti risultati.

Se le potenze j -esime ($j \neq 0$) dei termini della serie sono in progressione aritmetica il limite cui tende una media combinatoria

monoplana d'ordine c , o una media combinatoria biplana di ordini c , d , quando, fermi restando i termini estremi $x_1 = a$, $x_n = b$, il numero dei termini tende all'infinito, è dato da

$$\frac{b^{j+1} - a^{j+1}}{\left(\frac{1}{j} + 1\right) (b^j - a^j)}$$

Per una media monoplana combinatoria potenziata di ordini c , p , e per una media biplana combinatoria equipotenziata, d'ordini c , d , p , tale limite è dato da

$$\left[\frac{b^{p+1} - a^{p+1}}{\left(\frac{p}{j} + 1\right) (b^j - a^j)} \right]^{\frac{1}{p}} \quad \text{se } p \neq j$$

e da

$$\left[\frac{b^j - a^j}{j (\log b - \log a)} \right]^{\frac{1}{j}} \quad \text{se } p = -j$$

Per una media biplana equicombinatoria potenziata, d'ordini c , p , q , il limite da noi cercato è dato da

$$\left[\frac{(b^{p+j} - a^{p+j}) \left(\frac{q}{j} + 1\right)}{(b^{q+j} - a^{q+j}) \left(\frac{p}{j} + 1\right)} \right]^{\frac{1}{p-j}} \quad \text{per } p \neq -j \text{ e } q \neq -j$$

da

$$\left[\frac{j (b^{p+1} - a^{p+1}) (\log b - \log a)}{\frac{p}{j} + 1} \right]^{\frac{1}{p+j}} \quad \text{per } p \neq -j \text{ e } q \neq -j$$

e da

$$\left[\frac{\frac{q}{j} + 1}{j (b^{q+j} - a^{q+j}) (\log b - \log a)} \right]^{\frac{1}{-j-q}} \quad \text{per } p = -j \text{ e } q \neq -j$$

Infine per una media biplana combinatoria potenziata, d'ordini c, d, p, q , il limite da noi cercato è dato da

$$\left[\frac{\left\{ \frac{b^{p+i} - a^{p+i}}{\left(\frac{p}{j} + 1 \right) (b^j - a^j)} \right\}^c}{\frac{1}{c p - q d}} \right] \quad \text{per } p \neq -j \text{ e } q \neq -j$$

da

$$\left[\frac{\left\{ \frac{b^{p+i} - a^{p+i}}{\left(\frac{p}{j} + 1 \right) (b^j - a^j)} \right\}^c}{\frac{1}{c p + d}} \right] \quad \text{per } p \neq -j \text{ e } q = -j$$

e da

$$\left[\frac{\left\{ \frac{b^j - a^j}{j (\log b - \log a)} \right\}^c}{\frac{1}{-c - d q}} \right] \quad \text{per } p = -j \text{ e } q \neq -j$$

Resta ora da calcolare il medesimo limite per i vari tipi di medie combinatorie, nel caso che i termini siano *in progressione geometrica*. Si ottengono i seguenti risultati:

Per le medie combinatorie monoplane e biplane, il limite cercato è dato da

$$\frac{b - a}{\log b - \log a}$$

Per le medie monoplane combinatorie potenziate d'ordini c, p , e per le medie biplane combinatorie equipotenziate d'ordini c, d, p tale limite è dato da

$$\left[\frac{b^p - a^p}{p (\log b - \log a)} \right]^{\frac{1}{p}}$$

Per le medie biplane equicombinatorie potenziate d'ordini c, p, q , il limite cercato è dato da

$$\left[\frac{q (b^p - a^p)}{p (b^q - a^q)} \right]^{\frac{1}{p-q}}$$

Riportiamo ora i corrispondenti valori della media monoplana combinatoria potenziata con ripetizione, da noi calcolati.

TAVOLA II. — *Medie monoplane combinatorie, potenziate con ripetizione.*

$c \backslash p =$	- 4	- 3	- 2	- 1	0	1	2	3	4
1	1,47	1,64	1,99	3,01	8,33	26,60	45,83	58,65	66,90
2	1,39	1,49	1,83	2,73	8,33	30,60	54,49	63,77	71,30
3	1,34	1,46	1,72	2,52	8,33	34,42	55,77	67,43	74,37
4	1,30	1,41	1,64	2,36	8,33	37,82	59,18	70,24	76,69
5	1,27	1,35	1,56	2,19	8,33	40,02	63,12	72,85	78,41

I risultati ottenuti confermano pienamente, come si vede subito, quanto abbiamo dimostrato nel § 3.

Una seconda applicazione mirerà invece a mostrare praticamente quanto abbiamo asserito nel § 5, che cioè, se i termini di una serie sono in progressione aritmetica, quando, fissi restando gli estremi, il numero dei termini tende all'infinito, una qualunque loro media combinatoria tende alla media aritmetica degli estremi.

Costruiamo a tal uopo diverse serie i cui termini estremi siano sempre 1 e 33, e i cui termini siano in progressione aritmetica di ragione rispettivamente 32, 16, 8, 4, 2, 1. Avremo in sostanza le serie seguenti :

1°	1, 33
2°	1, 17, 33
3°	1, 9, 17, 25, 33
4°	1, 5, 9, 13, 17, 21, 25, 29, 33
5°	1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33
6°	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33

Calcoliamo, per ciascuna di dette serie, i valori della media combinatoria semplice e della media combinatoria con ripetizione d'ordine 2 (1). Si ottengono i risultati seguenti:

Numero d'ordine della serie. . .	1°	2°	3°	4°	5°	6°
Media combinatoria senza ripetizione	5,74	14,27	16,02	16,60	16,82	16,92
Media combinatoria con ripetizione	19,35	18,21	17,61	17,31	17,16	17,08

Come si vede subito osservando questa tavola, le due medie al crescere del numero dei termini, convergono rapidamente, la prima crescendo e la seconda decrescendo, verso la media aritmetica degli estremi, che è eguale a 17, in accordo a quanto abbiamo dimostrato per via teorica.

(1) Essendo i termini delle serie considerate diversi tra loro, le medie combinatorie di 1ª specie coincidono con quelle di 2ª specie.



ERNESTO PIZZETTI

Medie ascendenti e medie discendenti

1. Tutta una categoria di medie ponderate dei diversi tipi (aritmetiche, geometriche, armoniche, ecc.) può facilmente costruirsi, ricordando il procedimento di sommazione (c, k) delle serie dovuto a Cesàro e tanto noto in Matematica ⁽¹⁾: in tale procedimento ricorrono infatti alcune medie aritmetiche ponderate costruite secondo una legge ben determinata, sulla quale baseremo fondamentalmente le nostre considerazioni.

Incominciamo anzitutto col definire una categoria di medie aritmetiche ponderate che non sono altro (a meno di una differenza formale) che le medie di Cesàro propriamente dette: data pertanto una seriazione di termini positivi $a_0, a_1 \dots a_n$ disposti in ordine non decrescente di grandezza e maggiori od uguali all'unità positiva, costruiamo la somma del primo ordine:

$$s_n^{(1)} = a_0 + a_1 + \dots + a_n$$

e successivamente le somme degli ordini successivi:

$$s_n^{(2)} = s_0^{(1)} + s_1^{(1)} + \dots + s_n^{(1)}$$

$$\dots \dots \dots$$

$$s_n^{(k)} = s_0^{(k-1)} + s_1^{(k-1)} + \dots + s_n^{(k-1)}$$

(1) Per la teoria matematica del procedimento di Cesàro, cfr. E. CESÀRO, *Sur la multiplication des séries*, « Bulletin des Sciences Mathématiques » (2), 14-1890 pagg. 114-120 oppure G. VITALI e G. SANSONE, *Moderna teoria delle funzioni di variabile reale*. Parte II: G. SANSONE, *Sviluppi in serie di funzioni ortogonali* pagg. 77-93, Zanichelli, Bologna, 1935. Si può considerare come una applicazione del procedimento di Cesàro anche il metodo usato da Czuber e Pietra per la determinazione della differenza media: in proposito, cfr. E. CZUBER, *Beitrag zur Theorie statistischer Reihen*. « Versicherungswissenschaftlichen Mitteilungen », Neue Folge, 9 Band, 2 Heft, pagg. 118-123 e G. PIETRA, *The Theory of statistical relations with special reference to cyclical series*, « Metron », vol. IV, N. 3-4.

la *media armonica ascendente di ordine k* è data dalla formula

$$(4) \quad H_n^{(k)} = \frac{\binom{n+k}{k}}{l_n^{(k)}} = \frac{\binom{n+k}{k}}{\sum_{m=0}^n \frac{\binom{n-m+k-1}{k-1}}{a_m}}$$

La definizione, ora data nel caso delle medie aritmetiche, geometriche ed armoniche, di media ascendente, può facilmente generalizzarsi al caso di una media qualsiasi: intenderemo infatti come *media ascendente di ordine k* di un dato tipo dei termini a_0, a_1, \dots, a_n , la media di quello stesso tipo ponderata con peso $\binom{n-m+k-1}{k-1}$ per il termine a_m .

Alla categoria delle medie ascendenti fa riscontro un'altra categoria di medie, simmetriche alle precedenti e che denomineremo medie discendenti.

Così, nel caso delle medie aritmetiche, data la seriazione a_0, a_1, \dots, a_n ci costruiamo la seguente successione di somme del 1° ordine:

$$(1)s_0 = a_n \quad (1)s_1 = a_n + a_{n-1} \dots \quad (1)s_m = a_n + a_{n-1} + \dots + a_{n-m} \dots$$

$$(1)s_n = a_n + a_{n-1} + \dots + a_0$$

e quindi la seguente successione di somme del 2° ordine:

$$(2)s_0 = (1)s_0, \quad (2)s_1 = (1)s_0 + (1)s_1 \dots \quad (2)s_m = (1)s_0 + (1)s_1 + \dots + (1)s_m \dots$$

$$(2)s_n = (1)s_0 + (1)s_1 + \dots + (1)s_n$$

ed in generale la seguente successione di somme di ordine k :

$$\begin{aligned} (k)s_0 &= (k-1)s_0, \quad (k)s_1 = (k-1)s_0 + (k-1)s_1 \dots \\ \dots \quad (k)s_m &= (k-1)s_0 + (k-1)s_1 \dots + (k-1)s_m \dots \\ \dots \quad (k)s_n &= (k-1)s_0 + (k-1)s_1 + \dots + (k-1)s_n. \end{aligned}$$

La *media aritmetica discendente di ordine k* è allora data dalla formula:

$$(5) \quad {}^{(k)}C_n = \frac{{}^{(k)}s_n}{\binom{n+k}{k}} = \frac{\sum_{m=0}^n \binom{n-m+k-1}{k-1} a_{n-m}}{\binom{n+k}{k}} =$$

$$= \frac{\sum_{m=0}^n \binom{m+k-1}{k-1} a_m}{\binom{n+k}{k}}$$

Analogamente la *media geometrica discendente di ordine k* avrà per formula

$$(6) \quad {}^{(k)}G_n = \sqrt[n]{\prod_m a_{n-m}^{\binom{n-m+k-1}{k-1}}} = \sqrt[n]{\prod_m a_m^{\binom{n+k-1}{k-1}}}$$

e la *media armonica discendente di ordine k*

$$(7) \quad {}^{(k)}H_n = \frac{\binom{n+k}{k}}{\sum_{m=0}^n \frac{\binom{n-m+k-1}{k-1}}{a_{n-m}}} = \frac{\binom{n+k}{k}}{\sum_{m=0}^n \frac{\binom{m+k-1}{k-1}}{a_m}}$$

Generalizzando, intenderemo quindi come *media discendente di ordine k* di un dato tipo dei termini $a_0, a_1 \dots a_n$ la media ponderata del medesimo tipo, con peso $\binom{n-m+k-1}{k-1}$ per il termine a_{n-m} .

* * *

2. Il procedimento usato per dedurre le formule (2), (3), (4), (5), (6) delle medie ascendenti e delle medie discendenti implica naturalmente che il valore di k sia intero non negativo: tale limitazione si toglie però facilmente quando si ricorda che l'espressione $\binom{n-m+k-1}{k-1}$ non perde valore per qualsiasi valore positivo di k , intero o frazionario.

Vogliamo ora dimostrare alcune proprietà e dedurre alcune formule relative alle $C_n^{(k)}$ e ${}^{(k)}C_n$.

Anzitutto dall'espressione (1) di $C_n^{(k)}$, deduciamo subito la seguente formula

$$(8) \quad C_n^{(k)} = \frac{\binom{n+k-1}{k-1}}{\binom{n+k}{k}} C_n^{(k-1)} + \frac{\binom{n+k-2}{k-1}}{\binom{n+k}{k}} C_{n-1}^{(k-1)} + \dots +$$

$$+ \frac{\binom{k}{k-1}}{\binom{n+k}{k}} C_1^{(k-1)} + \frac{\binom{k-1}{k-1}}{\binom{n+k}{k}} C_0^{(k-1)}$$

che mette in luce la relazione che lega una media aritmetica ascendente alle medie aritmetiche ascendenti di ordine inferiore di una unità e con numero di termini non superiore ad n . La (8) ci serve inoltre per dimostrare la seguente proprietà fondamentale:

La successione delle $C_n^{(k)}$, per n variabile da zero all'infinito, è una successione monotona crescente, mentre la successione delle $C_n^{(k)}$, per k variabile da uno all'infinito, è una successione monotona decrescente che ha per limite il primo termine a_0 della seriazione.

Per dimostrare che prima parte della tesi, è sufficiente verificare che

$$(9) \quad C_n^{(k)} > C_{n-1}^{(k)}.$$

Perciò ricordando la (8), si ha:

$$\begin{aligned} C_n^{(k)} &= \frac{\binom{n+k-1}{k-1}}{\binom{n+k}{k}} a_0 + \frac{\binom{n+k-2}{k-1}}{\binom{n+k}{k}} a_1 + \dots + \\ &\quad + \frac{\binom{n-m+k-1}{k-1}}{\binom{n+k}{k}} a_m + \dots + \frac{\binom{k-1}{k-1}}{\binom{n+k}{k}} a_n \\ C_{n-1}^{(k)} &= \frac{\binom{n+k-2}{k-1}}{\binom{n+k-1}{k}} a_0 + \frac{\binom{n+k-3}{k-1}}{\binom{n+k-1}{k}} a_1 + \dots + \\ &\quad + \frac{\binom{n-m+k-2}{k-1}}{\binom{n+k-1}{k}} a_m + \dots + \frac{\binom{k-1}{k-1}}{\binom{n+k-1}{k}} a_{n-1} \end{aligned}$$

da cui si deduce, con calcoli elementari, che:

se $n > k$, il coefficiente di a_0 in $C_n^{(k)}$ è minore del coefficiente di a_0 in $C_{n-1}^{(k)}$ ed i coefficienti di tutti gli altri termini sono invece maggiori in $C_n^{(k)}$ che in $C_{n-1}^{(k)}$.

se $n > 2k$, i coefficienti di a_0 e di a_1 in $C_n^{(k)}$ sono minori dei corrispondenti in $C_{n-1}^{(k)}$, mentre i coefficienti di tutti gli altri termini sono invece maggiori in $C_n^{(k)}$ che in $C_{n-1}^{(k)}$.

.....

se $n > m k$, i coefficienti di a_0, a_1, \dots, a_{m-1} in $C_n^{(k)}$ sono minori dei corrispondenti in $C_{n-1}^{(k)}$ e i coefficienti di tutti gli altri termini sono invece maggiori in $C_n^{(k)}$ che in $C_{n-1}^{(k)}$.

Concludiamo quindi che, qualunque sia la relazione fra n e k , i termini maggiori della seriazione hanno peso relativo maggiore in $C_n^{(k)}$ che in $C_{n-1}^{(k)}$ e perciò che la (9) è senz'altro verificata: resta quindi dimostrato che la successione delle $C_n^{(k)}$, per n variabile da uno all'infinito, è una successione monotona crescente.

Sostituendo ora nella (8), al posto di $C_{n-1}^{(k-1)} \dots C_0^{(k-1)}$, $C_n^{(k)}$, eseguiamo una maggiorazione ed otteniamo pertanto:

$$C_n^{(k)} < \frac{\binom{n+k-1}{k-1} C_n^{(k-1)} + \binom{n+k-2}{k-1} C_{n-1}^{(k-1)} + \dots + \binom{n+k-1}{k-1} C_0^{(k-1)}}{\binom{n+k}{k}} +$$

$$+ \frac{\binom{k}{k-1} C_n^{(k-1)} + \binom{k-1}{k-1} C_n^{(k-2)}}{\binom{n+k}{k}} = C_n^{(k-1)} :$$

con ciò resta anche dimostrato che la successione delle $C_n^{(k)}$, al variare di k , è una successione monotona decrescente.

Calcoliamo infine il limite, a cui tende $C_n^{(k)}$ per $k \rightarrow \infty$. Perciò:

$$\lim_{k \rightarrow \infty} C_n^{(k)} = \lim_{k \rightarrow \infty} \frac{\sum_{m=0}^n \binom{n-m+k-1}{k-1} a_m}{\binom{n+k}{k}} =$$

$$= \sum_{m=0}^n \lim_{k \rightarrow \infty} \frac{\binom{n-m+k-1}{k-1}}{\binom{n+k}{k}} a_m =$$

$$= \sum_{m=0}^n \lim_{k \rightarrow \infty} \frac{k \cdot n \cdot (n-1) \dots (n-m+1)}{(n+k) \cdot (n+k-1) \dots (n+k-m)} a_m =$$

$$= \lim_{k \rightarrow \infty} \frac{k}{n+k} a_0 + \lim_{k \rightarrow \infty} \frac{n \cdot k}{(n+k)(n+k-1)} a_1 +$$

$$+ \dots + \lim_{k \rightarrow \infty} \frac{k \cdot n!}{(n+k) \dots k} a_n = a_0$$

Così la proprietà enunciata risulta completamente dimostrata.

Per le medie aritmetiche discendenti vale una proprietà del tutto simmetrica a quella che abbiamo ora dimostrata per le medie aritmetiche ascendenti e precisamente:

La successione delle ${}^{(k)}C_n$, per n variabile da zero all'infinito, rappresenta una successione monotona decrescente, mentre la medesima successione, per k variabile da uno all'infinito, rappresenta una successione monotona crescente che ha per limite l'ultimo termine a_n della seriazione.

È anzitutto opportuno osservare che mentre $C_{n-1}^{(k)}$ era una media relativa ai primi n termini della seriazione in esame e cioè ad $a_0, a_1 \dots a_{n-1}$, ${}^{(k)}C_{n-1}$ è invece una media relativa agli ultimi n termini della seriazione, $a_n, a_{n-1} \dots a_1$.

Con ragionamento analogo a quello fatto dianzi, si deduce immediatamente che ${}^{(k)}C_n < {}^{(k)}C_{n-1}$, in quanto i termini maggiori della seriazione hanno peso relativo maggiore in $C_{n-1}^{(k)}$ che in $C_n^{(k)}$, mentre i termini minori hanno peso relativo minore in $C_{n-1}^{(k)}$ che in $C_n^{(k)}$.

Poichè inoltre fra le ${}^{(k)}C_n$ e le ${}^{(k-1)}C_{n-1}, {}^{(k-1)}C_{n-2} \dots {}^{(k-1)}C_0$ vale una relazione identica alla (8), si ha subito

$${}^{(k)}C_n > \frac{\binom{n+k-1}{k-1} {}^{(k-1)}C_n + \binom{n+k-2}{k-1} {}^{(k-1)}C_{n-1} + \dots}{\binom{n+k}{k}} +$$

$$+ \frac{\binom{k}{k} {}^{(k-1)}C_0}{\binom{n+k}{k}} = {}^{(k-1)}C_n$$

Infine il limite di $C_n^{(k)}$, al tendere di k all'infinito, è:

$$l i m_{k \rightarrow \infty} {}^{(k)}C_n = \sum_{m=0}^n l i m_{k \rightarrow \infty} \frac{\binom{n-m+k-1}{k-1}}{\binom{n+k}{k}} a_{n-m} =$$

$$= l i m_{k \rightarrow \infty} \frac{k}{n+k} a_n + \dots + l i m_{k \rightarrow \infty} \frac{k \cdot n!}{(n+k) \dots k} = a_n.$$

Con ciò rimane completamente dimostrata la proprietà relativa alle ${}^{(k)}C_n$.

Particolare importanza assumono poi, le medie del secondo e del terzo ordine quando si vogliano mettere in evidenza alcune relazioni semplicissime fra le medie aritmetiche ascendenti e le medie aritmetiche discendenti.

Valgono infatti le seguenti proprietà :

$$a) \quad C_n^{(2)} + {}^{(2)}C_n = 2 C_n^{(1)} \quad (10)$$

poichè

$$\begin{aligned} C_n^{(2)} + {}^{(2)}C_n &= \frac{(n+1)a_0 + na_1 + \dots + a_n + (n+1)a_n + na_{n-1} + \dots + a_0}{\frac{(n+2)(n+1)}{2}} = \\ &= \frac{(n+2)a_0 + (n+2)a_1 + \dots + (n+2)a_n}{\frac{(n+2)(n+1)}{2}} = 2 C_n^{(1)}. \end{aligned}$$

cioè: la media aritmetica è la semisomma delle medie aritmetiche ascendenti e discendenti del secondo ordine.

$$b) \quad {}^{(2)}C_n - C_n^{(2)} = \frac{n}{n+2} \Delta, \quad (11)$$

ove Δ indica la differenza media semplice ⁽¹⁾ delle $(n+1)$ quantità a_0, a_1, \dots, a_n .

Infatti si ha :

$${}^{(2)}C_n - C_n^{(2)} = \frac{\sum_{m=0}^n (n-m+1) (a_{n-m} - a_m)}{\binom{n+2}{2}}$$

od anche

$${}^{(2)}C_n - C_n^{(2)} = \frac{\sum_{m=0}^{n/2} (n-2m) (a_{n-m} - a_m)}{\frac{(n+2)(n+1)}{2}}$$

(1) Per la definizione, le applicazioni della differenza media ed anche per tutta la letteratura al riguardo, cfr. C. GINI, *Memorie di Metodologia Statistica*. Vol. I: *Variabilità e Concentrazione*. II memoria: *Variabilità e Mutabilità*. Giuffrè, Milano, 1938.

D'altra parte, com'è noto la differenza media è data dalla formula (1)

$$\Delta = \frac{\sum_{m=0}^{n/2} (n-2m) (a_{n-m} - a_m)}{\frac{n(n+1)}{2}}$$

Confrontando le espressioni trovate, resta dimostrata la nostra formula (11), che non è altro (a meno di una differenza nei coefficienti numerici) che la formula di Czuber.

$$c) \quad {}^{(3)}C_n - C_n^{(3)} = \frac{3}{2} \cdot \frac{n}{n+2} \Delta \quad (12)$$

La dimostrazione di questa formula è del tutto analoga a quella della (11).

Dalla formula (12) si ha che la differenza media è uguale a ${}^{(3)}C_n - C_n^{(3)}$ quando è $n = 4$ e che per $n > 4$ è invece sempre $\Delta < ({}^{(3)}C_n - C_n^{(3)})$: pertanto, poichè dalla (11) si ha invece che, per qualsiasi valore di n , $\Delta > ({}^{(2)}C_n - C_n^{(2)})$, potremo dire che, quando è $n > 4$, la differenza media è compresa fra le due quantità ${}^{(3)}C_n - C_n^{(3)}$ e ${}^{(2)}C_n - C_n^{(2)}$.

Tutte le proprietà precedenti sono state dimostrate per le medie aritmetiche, ma è facile riconoscere che a proprietà analoghe ci si può ricondurre anche per le medie geometriche, qualora si tenga presente la seguente proprietà di carattere generale:

Il logaritmo della media geometrica ascendente (discendente) di ordine k dei termini a_0, a_1, \dots, a_n è uguale alla media aritmetica ascendente (o discendente) di ordine k dei logaritmi dei termini stessi.

Avremo quindi:

$$a') \quad G_n^{(2)} \cdot {}^{(2)}G_n = (G_n^{(1)})^2$$

$$b') \quad \log \frac{{}^{(2)}G_n}{G_n^{(2)}} = \frac{n}{n+2} \Delta.$$

Si verifica poi facilmente che tanto la successione delle $G_n^{(k)}$ che quella delle $H_n^{(k)}$, al variare di n , sono decrescenti, mentre quelle delle ${}^{(k)}G_n$ e ${}^{(k)}H_n$ sono crescenti.

(1) Cfr. C. GINI, l. c., pag. 208.

3. È interessante applicare le formule delle medie dianzi trovate al caso di alcuni particolari tipi di seriazione.

Supponiamo anzitutto che i termini a_0, a_1, \dots, a_n della seriazione che si considera siano i termini di una progressione aritmetica a ragione h :

$$a_0 = a, \quad a_1 = a + h, \quad a_2 = a + 2h \dots a_n = a + nh$$

e consideriamo le somme $s_n^{(k)}$ (naturalmente per valori interi di k):

$$s_n^{(1)} = (n+1)a + (1+2+\dots+n)h = (n+1)\left[a + \frac{nh}{2}\right]$$

$$\begin{aligned} s_n^{(2)} &= [1+2+\dots+(n+1)]a + \left[\binom{2}{2} + \binom{3}{2} + \dots + \binom{n+1}{2}\right]h = \\ &= \binom{n+2}{2}a + \binom{n+2}{3}h \end{aligned}$$

.....

$$\begin{aligned} s_n^{(k)} &= \sum_{m=0}^n \binom{n-m+k-1}{k-1} a_m = \sum_{m=0}^n \binom{n-m+k-1}{k-1} (a + mh) = \\ &= \binom{n+k}{k}a + \binom{n+k}{k+1}h \end{aligned}$$

Esaminando le espressioni ora trovate per le $s_n^{(k)}$, si nota subito che le $s_n^{(1)}$ sono termini di una progressione aritmetica del 2° ordine, le $s_n^{(2)}$ termini di una progressione aritmetica del 3° ordine ed in generale le $s_n^{(k)}$ di una progressione aritmetica di ordine $(k+1)$.

Le espressioni delle $C_n^{(k)}$ dedotte immediatamente dalle espressioni corrispondenti delle $s_n^{(k)}$ sono:

$$C_n^{(1)} = a + \frac{nh}{2}; \quad C_n^{(2)} = a + \frac{nh}{3}; \quad C_n^{(3)} = a + \frac{nh}{4}; \dots$$

$$\dots C_n^{\left(\frac{n}{2}-1\right)} = a + 2h; \dots C_n^{(k)} = a + \frac{nh}{k+1}; \quad C_n^{(n-1)} = a + h \dots$$

$$\dots C_n^{(2n-1)} = a + \frac{h}{2}; \dots$$

e mettono in evidenza le proprietà generali delle $C_n^{(k)}$ dimostrate al n. 2. È da notarsi inoltre che, se n è pari, $C_n^{(1)}$ coincide oltre

che con la media aritmetica, anche con la mediana, $C_n^{(4)}$ coincide con il primo quartile, $C_n^{(9)}$ con il primo decile e così via.

Analogamente alle $C_n^{(k)}$, si deducono le somme $^{(k)}S_n$, la cui formula generale, come si verifica elementarmente, è la seguente :

$$^{(k)}S_n = \binom{n+k}{k} a + \binom{n+k}{k} n h - \binom{n+k}{k+1} h = \binom{n+k}{k} a + \binom{n+k}{k+1} k h$$

dalla quale si deducono le medie aritmetiche discendenti $^{(k)}C_n$

$$^{(1)}C_n = a + \frac{n h}{2}; \quad ^{(2)}C_n = a + \frac{2}{3} n h; \quad ^{(3)}C_n = a + \frac{3}{4} n h \dots$$

$$^{(k)}C_n = a + \frac{k}{k+1} n h$$

Analogamente a quanto si verificava per le $C_n^{(k)}$; per n pari notiamo che $^{(3)}C_n$ coincide con il terzo quartile, $^{(9)}C_n$ con il nono decile e così via.

È infine da osservarsi che le medie $C_n^{(k)}$ e $^{(k)}C_n$ nel caso di una progressione aritmetica, oltre a verificare le proprietà generali già note, verificano anche per qualsiasi valore intero di k , la relazione

$$C_n^{(k)} + ^{(k)}C_n = 2 C_n^{(1)}$$

in quanto è evidentemente

$$C_n^{(k)} + ^{(k)}C_n = \frac{\binom{n+k}{k} a + \binom{n+k}{k+1} h + \binom{n+k}{k} a + \binom{n+k}{k+1} k h}{\binom{n+k}{k}} = 2 a + n h = 2 C_n^{(1)}$$

Supponiamo ora invece che i termini a_0, a_1, \dots, a_n si succedano in progressione geometrica a ragione q :

$$a_0 = a \quad a_1 = a q \quad a_2 = a q^2 \dots a_n = a q^n$$

e calcoliamo i prodotti dei vari ordini $p_n^{(k)}$ e $^{(k)}p_n$, prodotti che ci serviranno per dedurre le espressioni delle medie geometriche ascendenti e discendenti. Si ha precisamente :

$$\begin{aligned} p_n^{(1)} &= a^{n+1} q^{\frac{n(n+1)}{2}} & ^{(1)}p_n &= a^{n+1} q^{\frac{n(n+1)}{2}} \\ p_n^{(2)} &= a^{\binom{n+2}{2}} q^{\binom{n+2}{3}} & ^{(2)}p_n &= a^{\binom{n+2}{2}} q^{\binom{n+2}{3}} \\ &\dots\dots\dots & & \dots\dots\dots \\ p_n^{(k)} &= a^{\binom{n+k}{k}} q^{\binom{n+k}{k+1}} & ^{(k)}p_n &= a^{\binom{n+k}{k}} q^{\binom{n+k}{k+1}} \end{aligned}$$

Le espressioni delle $G_n^{(k)}$ e ${}^{(k)}G_n$ risultano pertanto le seguenti :

$$G_n^{(1)} = a q^{\frac{n}{2}} ; G_n^{(2)} = a q^{\frac{n}{3}} ; \dots G_n^{(k)} = a q^{\frac{n}{k+1}}$$

$${}^{(1)}G_n = a q^{\frac{n}{2}} ; {}^{(2)}G_n = a q^{\frac{2n}{3}} ; \dots {}^{(k)}G_n = a q^{\frac{kn}{k+1}}$$

4. *Applicazioni.* — Come verifica delle proprietà delle medie definite nei paragrafi precedenti, riportiamo i valori delle medie $C_n^{(k)}$, $G_n^{(k)}$, $H_n^{(k)}$, ${}^{(k)}C_n$, ${}^{(k)}G_n$, ${}^{(k)}H_n$ per alcuni particolari valori di k e per la seriazione a sette termini

$$a_0 = 1, a_1 = 2, a_2 = 4, a_3 = 5, a_4 = 8, a_5 = 10, a_6 = 20$$

Media	1° ordine	2° ordine	3° ordine	4° ordine	10° ordine
$C_6^{(k)}$	7,1429	4,3928	3,3095	2,7428	1,6894
$G_6^{(k)}$	4,8595	3,0655	2,4052	2,0673	1,4315
$H_6^{(k)}$	3,1460	2,1415	1,8523	1,6607	1,2838
${}^{(k)}C_6$	7,1428	9,8928	11,5595	12,7095	15,8817
${}^{(k)}G_6$	4,8595	7,7020	9,5760	10,9620	13,0840
${}^{(k)}H_6$	3,1460	5,4630	7,3850	8,8701	13,2680

Dai valori trovati per $C_n^{(2)}$ e ${}^{(2)}C_n$ sono verificate subito le formule (10), (11), (12): infatti la (10) è verificata in quanto: $C_6^{(2)} + {}^{(2)}C_6 = 4,3928 + 9,8928 = 14,2856 = 2 \times 7,1428 = 2C_6^{(1)}$ (10') la (11) in quanto

$${}^{(2)}C_6 - C_6^{(2)} = 9,8928 - 4,3928 = 5,5.$$

Ma la differenza media della seriazione considerata è:

$$\Delta = \frac{6(20 - 1) + 4(10 - 2) + 2(8 - 4)}{21} = \frac{154}{21} = 7,333$$

e quindi

$${}^{(2)}C_6 - C_6^{(2)} = 5,5 = \frac{3}{4} 7,333 \quad (11')$$

Infine la (12) è verificata poichè

$${}^{(3)}C_6 - C_6^{(3)} = 11,5595 - 3,3095 = 8,25 = \frac{3}{2} \cdot \frac{6}{8} 7,3333 = 8,25 \quad (12')$$

Per ultimo verifichiamo la formula relativa alle medie geometriche $(G_n^{(1)})^3 = G_n^{(2)} \cdot {}^{(2)}G_n$:

$$(4,8595)^3 = 23,6147 \quad 7,7020 \times 3,0655 = 23,6147$$

R. MOGNO

Su una formula approssimata per il calcolo di $n!$ e le sue applicazioni

SUNTO. — L'A. determina una formula più approssimata di quella di Stirling per il calcolo di $n!$ e l'applica per determinare una formula che consenta il calcolo dei valori delle probabilità di eventi rari con una precisione maggiore di quella consentita dalle altre formule comunemente in uso.

§ 1. — *Deduzione della formula.*

Ricordiamo la formula esatta del fattoriale di n :

$$n! = n^n \sqrt{2 \pi n} e^{-n + \frac{\delta}{12n}} \quad 1)$$

nella quale δ è un numero incognito compreso fra 0 ed 1.

Poichè δ non è noto, per calcolare il valore del fattoriale di n si usa la formula approssimata ⁽¹⁾ per difetto :

$$n! \simeq n^n \sqrt{2 \pi n} e^{-n} \quad 2)$$

Questa celebre formula, stabilita da De Moivre-Stirling, sostituisce alla funzione fattoriale, definita nei trattati elementari solo per valori interi e positivi dell'argomento, una funzione continua, sempre crescente per n compresa fra zero e $+\infty$.

L'errore assoluto che si commette applicando la 2) :

$$n^n e^{-n} \sqrt{2 \pi n} \left(e^{\frac{\delta}{12n}} - 1 \right) \quad 3)$$

(1) Per indicare che una uguaglianza è approssimata adoperiamo il segno : \simeq

cresce al crescere di n , mentre l'errore relativo :

$$\frac{e^{\frac{\delta}{12n}} - 1}{\frac{\delta}{12n}} \quad 4)$$

tende a zero al crescere di n .

Dalla formula di Wallis si deduce :

$$\lim_{n \rightarrow +\infty} 3 \cdot 5 \cdot 7 \dots (n-1) : \lim_{n \rightarrow +\infty} 2^{\frac{n}{2}} \left(1 \cdot 2 \cdot 3 \dots \frac{n}{2} \right) \left(\frac{n+1}{2} \cdot \pi \right)^{-\frac{1}{2}} = 1 \quad (n \text{ pari}) \quad 5)$$

Ora, poichè :

$$n! = 2^{\frac{n}{2}} \left(\frac{n}{2} ! \right) \left\{ 3 \cdot 5 \cdot 7 \dots (n-1) \right\} \quad 6)$$

da questa e dalla 5) si ricava :

$$\lim_{n \rightarrow +\infty} n! : \lim_{n \rightarrow +\infty} 2^n \left(\frac{n}{2} ! \right)^2 \cdot \left(\frac{n+1}{2} \cdot \pi \right)^{-\frac{1}{2}} = 1 \quad 7)$$

Dalla 1) si ha :

$$\lim_{n \rightarrow +\infty} n! : \lim_{n \rightarrow +\infty} n^n \sqrt{2\pi n} e^{-n} = 1 \quad 8)$$

Dividendo membro a membro la 7) per la 8) si ottiene :

$$\lim_{n \rightarrow +\infty} 2^n \left(\frac{n}{2} ! \right)^2 \left(\frac{n+1}{2} \pi \right)^{-\frac{1}{2}} : \lim_{n \rightarrow +\infty} n^n e^{-n} \sqrt{2\pi n} = 1 \quad 9)$$

dalla quale, dopo facili semplificazioni, si ha :

$$\lim_{n \rightarrow +\infty} n! : \lim_{n \rightarrow +\infty} n^{n + \frac{1}{4}} \cdot 2^{\frac{1}{4}} \cdot \pi^{\frac{1}{2}} (2n+1)^{\frac{1}{4}} e^{-n} = 1 \quad 10)$$

Moltiplicando membro a membro la 8) e la 10) ed estraendo le radici quadrate dei due prodotti si ottiene :

$$\lim_{n \rightarrow +\infty} n! : \lim_{n \rightarrow +\infty} n^{n + \frac{1}{8} + \frac{1}{8 \cdot 2^3}} \cdot e^{-n} \pi^{\frac{1}{2}} 2^{\frac{1}{3} + \frac{1}{3 \cdot 2^3}} (2n+1)^{\frac{1}{6} - \frac{1}{3 \cdot 2^3}} = 1 \quad 11)$$

Eseguendo le stesse operazioni fra la 10) e la 11) si ha :

$$\lim_{n \rightarrow +\infty} n! : \lim_{n \rightarrow +\infty} n^{n + \frac{1}{3} - \frac{1}{3 \cdot 2^4}} \cdot e^{-n} \cdot \pi^{\frac{1}{2}} \cdot 2^{\frac{1}{3} - \frac{1}{3 \cdot 2^4}} (2n + 1)^{\frac{1}{6} + \frac{1}{3 \cdot 2^4}} = 1 \quad 12)$$

e ripetendo lo stesso procedimento un numero k di volte, si ottiene, se k è pari :

$$\lim_{n \rightarrow +\infty} n! : \lim_{n \rightarrow +\infty} n^{n + \frac{1}{3} - \frac{1}{3 \cdot 2^{k+2}}} \cdot e^{-n} \cdot \pi^{\frac{1}{2}} \cdot 2^{\frac{1}{3} - \frac{1}{3 \cdot 2^{k+2}}} \cdot (2n + 1)^{\frac{1}{6} + \frac{1}{3 \cdot 2^{k+2}}} = 1 \quad 13)$$

e se k è dispari :

$$\lim_{n \rightarrow +\infty} n! : \lim_{n \rightarrow +\infty} n^{n + \frac{1}{3} + \frac{1}{3 \cdot 2^{k+2}}} \cdot e^{-n} \cdot \pi^{\frac{1}{2}} \cdot 2^{\frac{1}{3} + \frac{1}{3 \cdot 2^{k+2}}} \cdot (2n + 1)^{\frac{1}{6} - \frac{1}{3 \cdot 2^{k+2}}} = 1 \quad 14)$$

Per k tendente a $+\infty$ si ha :

$$\lim_{n \rightarrow +\infty} n! : \lim_{n \rightarrow +\infty} n^{n + \frac{1}{3}} e^{-n} \pi^{\frac{1}{2}} 2^{\frac{1}{3}} (2n + 1)^{\frac{1}{6}} = 1 \quad 15)$$

ossia :

$$\lim_{n \rightarrow +\infty} n! : \lim_{n \rightarrow +\infty} n^n e^{-n} \sqrt{2\pi n} \left(1 + \frac{1}{2n}\right)^{\frac{1}{6}} = 1 \quad 16)$$

dalla quale si ricava l'uguaglianza approssimata :

$$n! \simeq n^n e^{-n} \sqrt{2\pi n} \left(1 + \frac{1}{2n}\right)^{\frac{1}{6}} \quad 17)$$

formula tanto più approssimata quanto più grande è n .

§ 2. — Asintoticità della 17).

L'errore relativo che si commette applicando la 17) per calcolare $n!$ è :

$$\frac{e^{\frac{\delta}{12n}} - \left(1 + \frac{1}{2n}\right)^{\frac{1}{6}}}{e^{\frac{\delta}{12n}}} \quad 18)$$

Col tendere di n a $+\infty$ tale errore tende a zero, cioè la 17) tende asintoticamente a $n!$.

§ 3. — *Massimo valore dell'errore relativo 18).*

L'errore relativo 18), non può superare il valore assoluto dell'espressione :

$$1 - \frac{\left(1 + \frac{1}{2n}\right)^{\frac{1}{6}}}{1 + \frac{1}{11n}} \quad 19)$$

Infatti, poichè : $e^{\frac{1}{12n}}$ è il massimo valore che può assumere $e^{\frac{\delta}{12n}}$ e poichè è :

$$e^{\frac{1}{12n}} = 1 + \frac{1}{12n} + \frac{1}{2! (12n)^2} + \dots < 1 + \frac{1}{12n} + \frac{1}{(12n)^2} + \dots + \dots < 1 + \frac{1}{11n},$$

sostituendo nella 18) quest'ultimo valore al posto di $e^{\frac{\delta}{12n}}$ si ha l'espressione 19) che, per n costante, è il massimo valore che può assumere la 18).

§ 4. — *La formula 17) è approssimata per difetto.*

Basta dimostrare che il valore di essa è interno all'intervallo :

$$(n^n e^{-n} \sqrt{2\pi n} ; n!) \quad 20)$$

Sappiamo già che tale valore approssimato è maggiore dell'estremo inferiore dell'intervallo 20) ; dimostriamo ora che è minore di $n!$.

In quest'ultima ipotesi dovrebbe essere :

$$\frac{n!}{n^n e^{-n} \sqrt{2\pi n} \left(1 + \frac{1}{2n}\right)^{\frac{1}{6}}} :$$

$$\frac{(n+1)!}{(n+1)^{n+1} e^{-n-1} \sqrt{2\pi(n+1)} \left(1 + \frac{1}{2(n+1)}\right)^{\frac{1}{6}}} > 1 \quad 21)$$

mentre nell'ipotesi contraria, tale rapporto dovrebbe essere < 1 .

È facile mostrare che si verifica la 21); infatti dopo aver eseguito il rapporto e semplificato essa diventa:

$$\left(1 + \frac{1}{n}\right)^n \cdot \left(1 + \frac{1}{n}\right)^{\frac{1}{3}} \cdot \frac{1}{e} \cdot \left(\frac{2n+3}{2n+1}\right)^{\frac{1}{6}} > 1 \quad 22)$$

e poichè dalle disuguaglianze:

$$\frac{\left(1 + \frac{1}{n}\right)^n}{e} < 1; \left(1 + \frac{1}{n}\right)^{\frac{1}{3}} \cdot \left(\frac{2n+3}{2n+1}\right)^{\frac{1}{6}} > 1 \quad 23)$$

si deduce l'altra:

$$\frac{1}{e} > \frac{\left(1 + \frac{1}{n}\right)^n}{\left(1 + \frac{1}{n}\right)^{\frac{1}{3}} \cdot \left(\frac{2n+3}{2n+1}\right)^{\frac{1}{6}}} \quad 24)$$

sostituendo nella 22) il secondo membro di quest'ultima in luogo di $\frac{1}{e}$ si ottiene:

$$\frac{\left(1 + \frac{1}{n}\right)^{n+\frac{1}{3}} \cdot \left(\frac{2n+3}{2n+1}\right)^{\frac{1}{6}} \cdot \left(1 + \frac{1}{n}\right)^n}{\left(1 + \frac{1}{n}\right)^{\frac{1}{3}} \cdot \left(\frac{2n+3}{2n+1}\right)^{\frac{1}{6}}} = \left(1 + \frac{1}{n}\right)^{2n} > 1$$

disuguaglianza evidente che dimostra il sussistere della 22).

§ 5. — *Facciamo vedere sopra esempi numerici qualche applicazione della 17).*

Esattamente si ha: $5! = 120,00$

Con la 17) si ha: $5! \simeq 5^5 e^{-5} \sqrt{10\pi} \left(1 + \frac{1}{10}\right)^{\frac{1}{6}} = 119,91$

errore assoluto: $0,09$

errore relativo: $0,0008$

Con la formula di Stirling $5! = 118,10$
 errore relativo: $0,0160$
 cioè 20 volte maggiore.

Così, esattamente: $10! = 3.628.800$
 Con la 17) $10! = 3.628.100$
 errore assoluto: 700
 errore relativo: $0,0002$

Con la formula di Stirling: $10! = 3.598.699$
 errore relativo: $0,0083$
 cioè 41 volte maggiore.

Analogamente: $20! = 2.432.902 \times 10^{12}$
 Con la 17): $20! = 2.432.800 \times 10^{12}$
 errore assoluto: 102×10^{12}
 errore relativo: $0,000042$

Con la formula di Stirling: $20! = 2.422.786 \times 10^{12}$
 errore relativo: $0,004100$
 cioè 100 volte maggiore.

§ 6. — *Applicazioni della 17) per il calcolo della probabilità di un dato scarto in n prove ripetute, in casi di asimmetria.*

Nei casi di asimmetria molto accentuata, cioè in cui p è molto diverso da q , $p < q$, cioè di eventi rari, per calcolare P_l o probabilità di uno scarto $l = v - np$, si usa comunemente la formula di Poisson:

$$P_l \simeq \frac{(n p)^{np+l} e^{-np}}{(n p + l)!} \quad (26)$$

formula abbastanza semplice, benchè permanga in essa un fattoriale, ma di applicabilità molto limitata, come ha dimostrato il Surico ⁽¹⁾.

(1) Confronta: L. A. SURICO, «Giornale dell'Istituto italiano degli attuari». Anno II, N. 3.

Altra formula usata è la seguente presentata dal Castelnuovo nelle sue lezioni di Calcolo delle probabilità (1928-29):

$$P_l \simeq \frac{1}{\sqrt{2\pi q(n\phi + l)}} \left(\frac{n\phi}{n\phi + l} \right)^{n\phi + l} \cdot e^{l \cdot \left[1 - \frac{l}{2nq} \right]} \quad 27)$$

la quale, rispetto ai tre parametri ϕ , n , v , ha un campo di applicabilità molto più esteso della precedente e si può applicare con soddisfacenti risultati per ogni scarto:

$$l \leq \sqrt{\frac{3}{\phi}} \cdot \sqrt{n\phi q} \quad 28)$$

Tale formula, che rappresenta un notevole progresso sulla precedente per il calcolo di P_l nel caso di eventi rari, in corrispondenza agli scarti vicini ai massimi oltre i quali essa non è applicabile, dà valori che in certe questioni sono insufficientemente approssimati.

Vedremo ora come si possa ottenere un ulteriore progresso con una nuova *non complicata* formula ottenuta applicando la 17) e che permette il calcolo di P_l , anche nei casi suddetti, con approssimazione molto maggiore di quella che si otterrebbe applicando la 27).

Sostituendo ai fattoriali che compaiono nella:

$$P_l = \frac{n!}{(n\phi + l)!(nq - l)!} \phi^{n\phi + l} \cdot q^{nq - l} \quad 29)$$

i valori approssimati dati dalla 17), si ottiene:

$$P_l \simeq \frac{n^n e^{-n} \sqrt{2\pi n} \cdot \left(1 + \frac{1}{2n}\right)^{\frac{1}{6}} \phi^{n\phi + l} \cdot q^{nq - l}}{(n\phi + l)^{n\phi + l} \cdot (nq - l)^{nq - l} 2\pi e^{-n} \sqrt{(n\phi + l)(nq - l)} \left(1 + \frac{1}{2(n\phi + l)}\right)^{\frac{1}{6}} \cdot \left(1 + \frac{1}{2(nq - l)}\right)^{\frac{1}{6}}} \quad 30)$$

la quale, dopo facili semplificazioni, e dopo aver posto:

$$n\phi + l = v, \quad nq - l = n - v,$$

diventa:

$$P_l \simeq \frac{(2n + 1)^{\frac{1}{6}} \cdot n^{n + \frac{1}{3}} \phi^v q^{n - v}}{\sqrt{2^4 - \pi^3} v^{\frac{1}{3}} (n - v)^{n - v + \frac{1}{3}} \left(v + \frac{1}{2}\right)^{\frac{1}{6}} \cdot \left(n - v + \frac{1}{2}\right)^{\frac{1}{6}}} \quad 31)$$

Formula non soggetta ad alcuna restrizione sui valori di p, q, n, v e sempre applicabile tranne che quando $v = 0$ e $v = n$.

§ 7. — Carattere asintotico della 31).

La 31) tende al valore della formula esatta quando n tende a $+\infty$.

Infatti, poichè :

$$\lim_{n \rightarrow +\infty} \frac{n!}{n^n e^{-n} \sqrt{2\pi n} \left(1 + \frac{1}{2n}\right)^{\frac{1}{6}}} = 1$$

il rapporto tra la 29) e la 30) tende ad 1 quando n tende a $+\infty$ e quindi l'errore relativo che si commette applicando la 31) in luogo della 29) tende a zero.

§ 8. — Confine superiore dell'errore relativo che si commette applicando la 31) per calcolare P_1 .

Chiamiamo con r, r_1, r_2 , rispettivamente gli errori relativi che si hanno calcolando i valori di $n!, (np+l)!, (nq-l)!$ con la formula approssimata 17) e poniamo :

$$\begin{aligned} \varepsilon &= \frac{r}{1-r} = \frac{e^{\frac{\delta}{12n}}}{\left(1 + \frac{1}{2n}\right)^{\frac{1}{6}}} - 1 \\ \varepsilon_1 &= \frac{r_1}{1-r_1} = \frac{e^{\frac{\delta}{12(np+l)}}}{\left(1 + \frac{1}{2(np+l)}\right)^{\frac{1}{6}}} - 1 \\ \varepsilon_2 &= \frac{r_2}{1-r_2} = \frac{e^{\frac{\delta}{12(nq-l)}}}{\left(1 + \frac{1}{2(nq-l)}\right)^{\frac{1}{6}}} - 1 \end{aligned}$$

Si potrà scrivere esattamente :

$$n! = n^n e^{-n} \sqrt{2\pi n} \left(1 + \frac{1}{2n}\right)^{\frac{1}{6}} (1 + \varepsilon)$$

$$(n p + l)! = (n p + l)^{n p + l} e^{-n p - l} \sqrt{2 \pi (n p + l)} \left(1 + \frac{1}{2 (n p + l)} \right)^{\frac{1}{6}} (1 + \varepsilon_1)$$

$$(n q - l)! = (n q - l)^{n q - l} e^{-n q + l} \sqrt{2 \pi (n q - l)} \left(1 + \frac{1}{2 (n q - l)} \right)^{\frac{1}{6}} (1 + \varepsilon_2)$$

Sostituendo nella 29) ai fattoriali che vi compaiono le espressioni date dai secondi membri delle tre ultime uguaglianze, essa si traduce nella uguaglianza pure esatta :

$$P_l = \frac{1 + \varepsilon}{(1 + \varepsilon_1) (1 + \varepsilon_2)} P_l$$

nella quale P_l è l'espressione approssimata di P_l data dal secondo membro della 31).

L'errore relativo da cui è affetta la 31) sarà :

$$\mu = \frac{P_l - P_l}{P_l} = 1 - \frac{(1 + \varepsilon_1) (1 + \varepsilon_2)}{(1 + \varepsilon)} = \varepsilon - \varepsilon_1 - \varepsilon_2 < 0$$

Sarà perciò :

$$\mu > -\varepsilon_1 - \varepsilon_2$$

$$-\mu < \varepsilon_1 + \varepsilon_2$$

ossia :

$$-\mu < \frac{e^{\frac{\delta}{12(n p + l)}}}{\left(1 + \frac{1}{2 (n p + l)} \right)^{\frac{1}{6}}} + \frac{e^{\frac{\delta}{12(n q - l)}}}{\left(1 + \frac{1}{2 (n q - l)} \right)^{\frac{1}{6}}} - 2$$

Ricordando che :

$$e^{\frac{\delta}{12(n p + l)}} < 1 + \frac{1}{11 (n p + l)}$$

$$e^{\frac{\delta}{12(n q - l)}} < 1 + \frac{1}{11 (n q - l)}$$

si avrà :

$$-\mu < \frac{1 + \frac{1}{11 (n p + l)}}{\left(1 + \frac{1}{2 (n p + l)} \right)^{\frac{1}{6}}} + \frac{1 + \frac{1}{11 (n q - l)}}{\left(1 + \frac{1}{2 (n q - l)} \right)^{\frac{1}{6}}} - 2$$

Il massimo valore di $-\mu$ si ha quando $v = n p + l = 1$, $n = 2$, nel qual caso $-\mu < 0,039$.

Per $n = 10$, $v = 2$ si ha: $-\mu < 0,007$ e poichè in pratica conviene adoperare formule approssimate pel calcolo di P_l quando $v > 3$ ed $n \geq 40$, ne consegue che l'errore relativo che si commette applicando la 31) invece della formula esatta è sempre $< 0,01$, minore, cioè, di quello che si commette applicando la formula approssimata del Castelnuevo.

§ 9. — *Terminiamo con qualche applicazione pratica della 31).*

ESEMPIO 1^o. — Da un'urna contenente una palla bianca e quattro palle nere vengono eseguite $n = 100$ estrazioni, riponendo ogni volta la palla estratta nell'urna.

Qual'è la probabilità dello scarto $+10$ (30 palle bianche e 70 palle nere)?

Con la formula esatta si ha:

$$P_{+10} = 0,00518966$$

Con la formula del Castelnuevo:

$$P_{+10} = 0,00500 \text{ con l'errore relativo} = 0,038$$

Ponendo nella 31)

$$n = 100, \quad p = \frac{1}{5}, \quad q = \frac{4}{5},$$

essa diventa:

$$P_l \simeq \frac{201^{\frac{1}{6}} \cdot 100^{100 + \frac{1}{3}} \cdot \left(\frac{1}{5}\right)^{30} \cdot \left(\frac{4}{5}\right)^{70}}{\sqrt{16 \pi^3} \cdot 30^{30 + \frac{1}{3}} \cdot 70^{70 + \frac{1}{3}} \cdot 30,5^{\frac{1}{6}} \cdot 70,5^{\frac{1}{6}}} = 0,00518982$$

con un errore relativo di: 0,000031 cioè 1000 volte più piccolo del precedente.

Un risultato tanto approssimato potrebbe solo ottenersi sviluppando P_l in serie di polinomi G del Jordan, ma con calcoli molto più lunghi e laboriosi di quelli richiesti dalla 31).

ESEMPIO 2°. — Facciamo ora un caso di asimmetria fortissima.

Sia ad es. $p = \frac{1}{1000}$; $q = \frac{999}{1000}$; $n = 100$ e si domandi la probabilità che l'evento di probabilità p , si presenti 1 volta in 100 prove ripetute (estrazioni).

Esattamente si ha:

$$P_1 = 0,0906$$

Con la formula del Castelnovo:

$$P_1 = 0,0974 \text{ con un errore relativo: } 0,075$$

Con la 31) si ottiene:

$$P_1 = 0,0918$$

con l'errore relativo: 0,013.

ESEMPIO 3°. — Sia ora $p = \frac{1}{4}$, $q = \frac{3}{4}$, $n = 32$ e si chieda la probabilità che l'evento di probabilità p si presenti 16 volte.

Esattamente si ha:

$$P_{+8} = 0,001403$$

Con la 31): $P_{+8} = 0,001403$; con la formula del

Castelnovo: $P_{+8} = 0,001398$; con la formula di

Charlier: $P_{+8} = 0,001411$; con la formula asimmetrica

di Laplace: $P_{+8} = 0,001448$; con la formula di

Poisson: $P_{+8} = 0,004500$

Dagli esempi addotti e dai confronti fatti si conclude:

1°) Con la nuova formula si ottengono risultati molto più approssimati di quelli ottenuti con le altre formule approssimate più comunemente in uso.

2°) L'applicazione della nuova formula approssimata non richiede calcoli più complicati di quelli richiesti dalla formula del Castelnovo o da altre formule o sviluppi in serie.

3°) Quando è necessario calcolare P_1 con grande approssimazione, la 31) è la formula più conveniente.

G. DE MEO

Sulle nascite dei primogeniti in Italia

I.

§ 1. — DIFFICOLTÀ PER LA MISURA DELLA FECONDITÀ IN GENERALE.

È noto che i comuni quozienti specifici di natalità $x_i = \frac{n_i}{m_i}$ dove con n_i si indicano i partoriti in un anno dalle m_i coniugate in età i , rappresentano in maniera piuttosto grossolana la probabilità per una coniugata dell'età i di divenire madre entro l'anno. Difatti, come ha fatto rilevare il Gini ⁽¹⁾, vi sono varie circostanze che tendono a far perdere molto valore a questi rapporti:

1) *Concepimenti antenuziali.* — Le nascite che si producono nei primi mesi del matrimonio, rappresentano dei casi reali che non trovano adeguata corrispondenza nei casi possibili. Difatti, fra le coppie che hanno rapporti intimi prima del matrimonio, legalizzano subito la loro unione col matrimonio, con maggiore probabilità, quelle che hanno anche dato luogo ad un concepimento. La correzione consistente nell'eliminare i parti verificatisi entro i primi nove mesi del matrimonio sia dalle nascite sia dalle donne maritate, non produce l'effetto desiderato, in quanto, si eliminano

(1) C. GINI, *Sur la mesure de la fécondité des mariages.* « Bulletin de l'Institut International de Statistique », Tome XXVII, La Haye, 1934; *Sulle relazioni tra le oscillazioni mensili del numero dei matrimoni e quelle del numero delle nascite*, « Metron », Vol. X, nn. 1-2; *Su la determinazione dei quozienti di eliminazione e in particolare sui metodi delle durate esatte e delle durate medie nell'ipotesi di saggi istantanei di eliminazione costanti*, « Metron », Vol. XII, n. 3, 1935; *Di un procedimento per la determinazione del numero medio di figli per matrimonio*, « Metron », Vol. X, nn. 1-2, 1932.

dal numeratore tutti i casi reali ma non si eliminano dal denominatore tutti i casi possibili, cioè a dire le donne che hanno avuto rapporti prenzuziali e che avrebbero potuto essere fecondate, ma non lo sono state.

2) *Impossibilità di avere figli entro i primi mesi del matrimonio.* — Entro questo periodo, la donna non rappresenta un caso possibile, e questa circostanza tende ad aumentare il denominatore del rapporto, e perciò tende a dare una misura approssimata per difetto della fecondità.

3) *Impossibilità di concepire subito dopo un parto.* — Durante vari mesi dopo un parto, per il ritardo nella ricomparsa delle mestruazioni che si accompagna spesso all'allattamento, la donna non rappresenta un caso possibile. Il denominatore della frazione risulta indebitamente accresciuto, e perciò indebitamente diminuito il quoziente di fecondità.

A questi inconvenienti dei comuni quozienti di fecondità, se ne potrebbero aggiungere altri:

a) In particolare, come osserva il Savorgnan ⁽¹⁾, il numero delle coniugate è composto di due elementi eterogenei rispetto alla fertilità: vi sono donne che non hanno mai partorito e donne che hanno già partorito. Ora, la probabilità di maternità secondo le vecchie indagini del Körösi e di altri ⁽²⁾ è molto maggiore per le donne che non hanno mai partorito, sia perchè nei matrimoni con prole spesso si tende a limitare i concepimenti, e sia perchè la donna può già avere esaurito il suo potere genetico per i precedenti concepimenti. Inoltre, poichè nel periodo di allattamento spesso manca la mestruazione, accade che la madre che allatta, è meno esposta ad un nuovo concepimento. La probabilità di maternità per le coniugate dipende anche dall'epoca alla quale risale il loro matrimonio, e dal diverso numero di volte che hanno già partorito. Così una donna di 30 anni che sia sposata da tre anni, ed abbia partorito una volta sola, è molto più esposta all'evento della maternità di una sua coetanea sposata da 10 anni e che abbia già dato alla luce 4 bambini.

b) Un'altra causa che tende a fare apparire più forte del vero la diminuzione della capacità genetica nel corso dell'età, desunta dai suddetti quozienti specifici, è data dall'azione selet-

(1) F. SAVORGNAN, *Corso di Demografia*, Pisa, 1936, pag. 121.

(2) R. BENINI, *Principi di Demografia*, Firenze, 1901, pag. 64 e segg.

tiva del matrimonio. Difatti, chi si unisce in matrimonio, si trova normalmente in condizioni di salute ottime ed è per lo più sicuro di poter convivere di fatto con la sposa. Quanto più si allontana la data del matrimonio, tanto più è probabile l'insorgere di cause sanitarie, sociali e morali che portano ad una convivenza inattiva o ad una separazione personale ⁽¹⁾.

Da quanto abbiamo detto, risulta evidente, che quando si vuol paragonare la fecondità di un gruppo con quella di un altro gruppo — anche se si elimina la influenza della composizione per età — il confronto è sempre in notevole misura alterato dalle circostanze di cui sopra.

§ 2. — LA FECONDITÀ DELLE PRIMIPARE.

Date queste difficoltà che rendono difficili e forse praticamente impossibili confronti rigorosi, si spiega come da tempo gli statistici — e fra questi, fra i primi, il Gini ⁽²⁾ ed il Savorgnan ⁽³⁾ — abbiano portato la loro attenzione sulle sole nascite dei primogeniti. Lo studio della frequenza delle nascite dei primogeniti da un gruppo di matrimoni e la distribuzione di queste nascite nel tempo, è infatti di grande importanza, in quanto, i risultati che si ottengono — se pure danno un'idea soltanto parziale del fenomeno della fecondità di una popolazione — possono ritenersi in gran parte esenti dagli inconvenienti di cui si è fatto cenno. Difatti:

1) Il gruppo delle donne studiate è uniforme per ciò che riguarda la probabilità di maternità derivante dal non aver mai partorito, salvo casi eccezionali.

(1) L. LIVI, *Lezioni di demografia*, Padova, 1936, pag. 56.

(2) C. GINI, *Prime ricerche sulla fecondabilità della donna*, in: « Atti del R. Istituto Veneto di Scienze, Lettere ed Arti », Tomo LXXXIII, Parte II, pagg. 315 e segg.; *Nuove ricerche sulla fecondabilità della donna*, ivi, Tomo LXXXIV, Parte II, pagg. 269 e segg.; *Le basi scientifiche della politica della popolazione*. Catania, 1931, pagg. 180 e segg.; *Nascita Evoluzione e Morte delle Nazioni*. Roma 1930, pagg. 58 e segg.; *Sur la mesure de la fécondité des mariages*, cit.; *La fecondità della donna in relazione al problema dell'eugenica*, in « Economia », agosto-settembre 1927; *Problemi della popolazione*, in: « Annali di statistica della R. Università di Bari », 1928.

(3) F. SAVORGNAN, *Le statistiche delle nascite secondo l'ordine di generazione*, Roma, 1933, XI; *Nuzialità e fecondità delle case Sovrane d'Europa*, « Metron », Vol. III, n. 2, 1923; *La fecondità delle aristocrazie*, « Metron », Vol. III, nn. 3-4, 1924 e Vol. IV, nn. 3-4, 1925.

2) Il gruppo è anche abbastanza uniforme sotto l'aspetto delle condizioni fisiche che possono favorire la maternità, perchè nei primi tempi del matrimonio, quando cioè si verificano la maggioranza delle nascite dei primogeniti, è ancora notevole l'effetto della selezione che si opera col matrimonio.

3) Poichè il desiderio di limitare o ritardare le nascite si inizia in generale dopo la nascita del primogenito, le eventuali variazioni nel tempo e nello spazio delle distribuzioni dei primogeniti, possono, in notevole misura, riconnettersi a fattori di indole biologica. È vero che le dette variazioni possono attribuirsi anche al fatto che in alcuni gruppi di popolazione possono esistere molte coppie che ritardano la nascita del primogenito per continuare a godere in libertà la vita matrimoniale, ma non è ragionevole supporre che tali coppie desiderino rimandare quella nascita oltre il 2° o 3° anno di matrimonio come risulta da alcune ricerche del Gini ⁽¹⁾ e come mostreranno le nostre elaborazioni (cfr. § 9). D'altra parte, si può osservare che, anche ammessa l'esistenza di queste coppie che intendono ritardare la prima nascita, non si può ragionevolmente ammettere che la loro percentuale sia molto elevata, specie in Italia, paese di media natalità. Nè si deve dimenticare che, poichè per le nuove spose la facilità al concepimento sarebbe notevolmente maggiore di quella relativa alle coetanee che già hanno dato luogo a figliolanza ⁽²⁾, si deve am-

(1) Cfr. nota 2 a pag. 81.

(2) Il GINI (C. GINI, *Nuove ricerche*, ecc. cit., pag. 288) esamina singolarmente le varie cause che tendono a far risultare differente la *fecondabilità* (cioè la probabilità di concepimento in un periodo mestruale all'infuori di ogni limitazione volontaria della prole) delle donne maritate in età riproduttiva dalla fecondabilità relativa alle donne in età riproduttiva che non hanno ancora avuto figli. Queste circostanze sono:

a) vizi di conformazione che talvolta possono venire corretti con successivi interventi ostetrici;

b) stenosi dei canali genitali che scompaiono dopo il primo parto;

c) strascichi delle gravidanze che spesso riducono o annullano la fecondabilità della donna;

d) esaurimento della facoltà riproduttiva della donna per successive gravidanze;

e) diminuita frequenza dei contatti sessuali per le donne che hanno avuto già figli, per le frequenti assenze del marito, per l'attenuarsi degli stimoli sessuali, per l'aumentata età, per l'intiepidirsi dell'affetto, ecc.

Le circostanze a) e b) tendono ad elevare la fecondabilità delle donne che hanno già avuto figli; le circostanze c), d), e) tendono ad abbassarla ed hanno la prevalenza sulle prime.

mettere che le pratiche contraccettive messe in atto dalle nuove coppie, riescano nel loro scopo in un numero di casi inferiore a quello certo non elevato ⁽¹⁾ che si riscontra per le spose in generale.

II.

Quando si disponga di una distribuzione di primogeniti in funzione dell'intervallo di tempo trascorso fra il matrimonio e la nascita, si può studiare la distribuzione stessa con vari metodi che esamineremo brevemente prima di illustrare i risultati delle nostre elaborazioni.

§ 3. — L'INTERVALLO PROTOGENESICO.

Fin dal 1917 il Knibbs ²⁾ aveva richiamato l'attenzione degli studiosi sull'intervallo medio fra il matrimonio e la prima nascita, avendo escluso le nascite registrate prima dei 9 mesi dal matrimonio. Questo intervallo medio fu detto dallo stesso Knibbs, *intervallo protogenesico*, e fu considerato dallo stesso A. come uno dei migliori indici misuratori dell'elemento fisiologico della fecondità della donna.

Ora, questo indice, come nota il Gini ⁽³⁾, si presta bene a misurare la fecondità fisiologica della donna solo in alcuni casi, mentre in altri, esso porta a conclusioni contrastanti. Così accade, ad esempio, quando si vuole desumere dall'intervallo protogenesico la fecondità della primipara in funzione dell'età *al matrimonio*. Poichè infatti, in generale l'*I_p* tende a diminuire, dopo una certa età, con l'aumento dell'età al matrimonio, vi sarebbe da concludere che aumentando l'età al matrimonio vada aumentando anche la facilità al concepimento, cosa che è evidentemente assurda. Ma il risultato si spiega considerando che fra le donne che si spo-

(1) Il GINI (*Nascita evoluzione e morte delle Nazioni*, cit. pag. 58) osserva, in base ai risultati di alcune statistiche americane, che le pratiche contraccettive nella popolazione generale non raggiungono il loro effetto in circa 2/3 dei casi.

(2) KNIBBS, *The mathematical Theory of Population*, Melbourne, 1917, pagg. 268-272.

(3) C. GINI, *Sur la mesure de la fécondité des mariages*, cit. pag. 41.

sano in età avanzata, per la prossimità dell'età della menopausa, alcune soltanto — le più feconde — riescono a concepire nel breve tempo disponibile, mentre tutte le altre escono dall'osservazione dopo la scomparsa delle mestruazioni ⁽¹⁾.

Studiando l'intervallo protogenesico in funzione dell'età *al parto* si perviene anche a risultati che non possono assumersi come del tutto significativi. Difatti, come fu rilevato dallo Knibbs ⁽²⁾ e come risulta dalle nostre elaborazioni, l'intervallo protogenesico è regolarmente crescente col crescere dell'età della madre al parto. Ora, ciò si verifica, almeno in parte, perchè, a parità di condizioni, più l'intervallo è lungo, e maggiore risulta l'età della madre.

Malgrado questi inconvenienti, non vi è dubbio che in alcuni casi si possa convenientemente usare l'*Ip* per avere una misura della fecondabilità delle primipare. Così accade, ad esempio per le spose classificate in funzione del mese di matrimonio, perchè si può ammettere che da un mese all'altro, non vari l'influenza dei fattori perturbanti ⁽³⁾. Così ancora, quando si possa ammettere che in due differenti gruppi di popolazione l'età media alla menopausa non sia apprezzabilmente differente, si può desumere dall'intervallo protogenesico una conveniente misura della fecondità delle primipare in quei due gruppi.

L'uso dell'intervallo protogenesico come misura della fecondità delle primipare può essere esteso, con alcuni accorgimenti, anche all'esame della fecondità delle primipare in funzione dell'età al matrimonio. Se, difatti, si considerano soltanto le donne che al matrimonio avevano un'età compresa fra i 15 e i 34 anni, e si effettua il calcolo dell'*Ip* utilizzando solo le nascite di primogeniti verificatesi al di sotto dei 10 anni dal matrimonio (che rappresentano poi la grandissima maggioranza delle nascite di primogeniti), allora i risultati ottenuti per le varie età, sono abbastanza bene paragonabili fra loro. Difatti, tanto le donne sposate a 15 anni, quanto quelle sposate a 34 anni, hanno lo stesso periodo fecondo innanzi ad esse (10 anni) per procreare il primogenito. È vero che

(1) KNIBBS, op. cit. pag. 272, spiegò l'apparente incongruenza dei risultati. Egli dice infatti: « It is evident that it must necessarily have a small value at the end of the child-bearing period ».

(2) KNIBBS, op. cit., pag. 270.

(3) C. GINI, *Sur la mesure*, ecc. cit., pag. 42.

fra le donne sposate a 34 anni, rispetto a quelle sposate a 15, è più frequente la scomparsa delle mestruazioni entro i 10 anni successivi, ma ciò deve considerarsi comunque un fatto anormale; e d'altra parte, un'elevata frequenza di menopause precoci fra le donne sposatesi in età avanzata, può, in un certo senso, considerarsi anche un'indice della loro bassa fecondabilità.

Se si suppone che la fine del periodo fecondo possa fissarsi a 49 anni, in luogo di considerare soltanto le donne di età compresa fra 15 e 34 anni, si possono considerare le donne di età compresa fra 15 e 39 anni ($39 + 10 = 49$). In tal caso, gli intervalli medi per le varie età, sono tanto meglio paragonabili fra loro, per quanto più è attendibile la detta ipotesi.

§ 4. — LA FECONDABILITÀ.

Nel 1924 il Gini ⁽¹⁾ introdusse il concetto di *fecondabilità*. Dicesi « fecondabilità della donna la probabilità che, in regime matrimoniale, all'infuori di ogni pratica contraccettiva e di ogni astensione dai contatti sessuali diretta a limitare la procreazione, la donna venga fecondata durante un periodo mestruale » ⁽²⁾. Siano n spose fecondabili aventi tutte la stessa fecondabilità; se p' è la fecondabilità in un dato periodo mestruale, e p'' la fecondabilità nel periodo successivo 2 di quelle spose del gruppo che non sono restate fecondate nel periodo 1, avremo:

n = numero iniziale delle donne fecondabili;

$1 - p'$ = probabilità per una sposa di non esser fecondata nel periodo 1;

$n(1 - p')$ = numero delle donne non fecondate nel periodo 1 esposte ad essere fecondate nel periodo 2;

$n(1 - p') p''$ = numero di donne fecondate nel periodo 2;

$n p'$ = numero di donne fecondate nel periodo 1.

Il rapporto fra il numero delle donne fecondate nel periodo 2 al numero delle donne fecondate nel periodo 1, sarà:

$$\frac{n(1 - p') p''}{n p'} = \frac{(1 - p') p''}{p'} \quad [1]$$

(1) C. GINI, *Prime ricerche*, ecc., cit.; *Nuove ricerche*, ecc., cit.

(2) C. GINI, *Nuove ricerche*, ecc., cit., pag. 272.

Qualora si possa ammettere che la fecondabilità delle donne fecondabili in un periodo, non sia sensibilmente diversa da quella delle donne fecondabili nel periodo successivo, supponendo cioè $p' = p''$, la [1] diviene:

$$1 - p'. \quad [2]$$

Ciò significa che il rapporto del numero delle spose fecondate nel periodo 2 al numero delle spose fecondate nel periodo 1, ammesso che il numero delle osservazioni sia sufficientemente grande, fornisce il complemento all'unità della misura della fecondabilità delle spose fecondabili nel periodo 1 ⁽¹⁾.

Il fatto che vi sia un certo numero r di spose del gruppo che si astengono dai contatti sessuali o si danno a pratiche contraccettive, non avrà alcuna influenza sul valore del rapporto 1 purchè tale numero sia stato uguale nei due periodi e le pratiche messe in uso, abbiano in ogni caso impedito la fecondazione. In tal caso, infatti, al posto di n al numeratore e al denominatore, andrebbe sostituito $n - r$, ma il valore del rapporto, rimarrebbe inalterato.

Non è praticamente possibile conoscere il numero delle spose fecondate nei periodi 1 e 2, ma se si ammette che per i due periodi successivi risulti eguale la frequenza:

a) delle interruzioni di gravidanza nei primissimi tempi (aborti ovulari);

b) degli aborti e dei nati morti;

c) dei parti prematuri (7-8 mesi) o tardivi (10 mesi);

d) dei nati vivi a termine;

il valore:

$$1 - \frac{\text{Nati vivi tra 9 mesi} + 27 \text{ giorni} + 9 \text{ mesi} + 54 \text{ giorni}}{\text{Nati vivi tra 9 mesi} + 9 \text{ mesi} + 27 \text{ giorni}} = p'$$

darà la misura della fecondabilità nel periodo 1. In pratica, poichè le statistiche non danno il numero dei nati nei vari periodi mestruali, il valore suddetto viene calcolato rapportando le nascite registrate fra il 10^o e l'11^o mese alle nascite fra il 9^o e il 10^o mese dal matrimonio. Si ottiene così un valore approssimato per eccesso, che andrebbe diminuito del 10% per tener conto approssimativamente della differenza fra il mese di calendario ed il periodo mestruale.

(1) C. GINI, *Nuove ricerche*, ecc., cit., pag. 282.

Il calcolo della fecondabilità implica varie ipotesi ⁽¹⁾ le quali però o possono ammettersi senz'altro, o (fatta eccezione per la determinazione della fecondabilità del 1° mese) non sembrano poter esercitare una influenza sensibile sui risultati. Fra tali ipotesi, una delle principali è la seguente: che le donne maritate fecondabili in un certo mese, abbiano la stessa fecondabilità di quelle fecondabili nel mese seguente. Ora, come è ovvio, questa ipotesi può essere lecita, qualunque sia l'età al matrimonio, quando ci si limita a considerare i due o tre primi mesi del matrimonio ⁽²⁾.

(1) Riportiamo qui appresso dal GINI (*Nuove ricerche, ecc., cit.*, pagg. 284-286) le ipotesi, oltre quella di un sufficiente numero di osservazioni, su cui si fonda il calcolo della fecondabilità:

a) sia trascurabile il movimento naturale e sociale nel gruppo delle spose considerate;

b) la fecondabilità in un periodo non sia sensibilmente diversa da quella del periodo successivo;

c) che il numero delle spose, che si astengono dai contatti sessuali o si danno a pratiche contraccettive, sia lo stesso nei due periodi;

d) che queste pratiche dirette ad impedire la fecondazione, raggiungano sempre il loro effetto;

e) che la frequenza delle spose che si danno a pratiche dirette ad evitare la fecondazione sia la stessa per i vari gruppi di spose aventi diversa fecondabilità;

f) che la frequenza delle interruzioni di gravidanza nei primissimi tempi (aborti ovulari) e degli aborti e nati morti, sia la stessa per i concepimenti dei due periodi;

g) che la frequenza delle gestazioni di durata inferiore e di quelle di durata superiore ai 9 mesi, sia la stessa pei concepiti nei 2 periodi;

h) che i nati vivi a termine rappresentino la stessa percentuale tra i nati vivi dei successivi periodi, a partire da nove mesi dal matrimonio.

Le ipotesi a), b), c), f), g), h), possono senz'altro essere ammesse. Vi è solo da notare a questo riguardo che la frequente condizione di verginità al matrimonio, fa sì che non si verifichi l'ipotesi b) per il 1° e 2° periodo mestruale dal matrimonio; ed analogamente che l'ipotesi f) non sembra valida per gli stessi periodi a causa dei viaggi di nozze che tenderebbero ad aumentare la frequenza degli aborti ovulari. Le ipotesi d) ed e), invece, non rispondono al vero: ma la loro influenza deve essere piuttosto limitata, perchè, nella maggioranza delle popolazioni, secondo il Gini, le pratiche dirette ad evitare la fecondazione, sono raramente impiegate dalle spose che non hanno ancora avuto figli.

(2) Sta qui il principale vantaggio della determinazione della fecondabilità su quella dell'*Ip* per la quale non si potrebbe adottare siffatta ipotesi, ed in generale non si può trascurare la circostanza che la fine del periodo riproduttivo è più vicina per le donne maritate ad un'età più avanzata. (Cfr. C. GINI, *Sur la mesure, ecc. cit.*, pag. 41-42).

Una volta ottenuti i valori di p per le varie distanze dal matrimonio, è evidente che si può calcolare la fecondabilità media del gruppo studiato, facendo la media dei p , ponderata in base al rispettivo numero di nascite verificatesi ai vari intervalli ⁽¹⁾.

Per la determinazione della fecondabilità p occorre tener presenti particolari accorgimenti quando le statistiche dei nati, anzichè fornire i dati mese per mese, danno il numero dei nati in periodi successivi più lunghi di un mese.

Nel caso che i successivi periodi abbiano tutti la stessa durata di n mesi, la formula che determina il valore di p , è:

$$N_2/N_1 = (1 - p)^n \quad [3]$$

dove N_1 è il numero dei nati in un certo periodo di n mesi e N_2 il numero dei nati nel periodo successivo.

Se i due periodi hanno durata diversa di m mesi il primo e di n mesi il secondo, $n = mt$ essendo un multiplo di m , il valore di p sarà dato dalla formula:

$$\frac{N_n}{N_m} = (1 - p)^m + (1 - p)^{2m} + \dots + (1 - p)^{tm} \quad [4]$$

dove N_m sono i nati nel primo periodo ed N_n i nati nel secondo ⁽²⁾. Tali formule, come avverte il Gini ⁽³⁾, conducono a risultati approssimati, ma pur sufficienti a dare un'idea del fenomeno.

§ 5. — DISTRIBUZIONI PERCENTUALI.

La determinazione dell'intervallo protogenesico, può essere in certi casi vantaggiosamente sostituita o integrata dall'esame della distribuzione percentuale delle prime nascite in funzione dell'intervallo di tempo trascorso dal matrimonio. In tal modo, l'esame dettagliato delle varie distribuzioni, può permettere di scorgere delle particolarità che, al contrario, sfuggirebbero in base alla sola comparazione delle medie.

§ 6. — CONCEPIMENTI NEI PRIMI TRE MESI DEL MATRIMONIO.

In alcuni casi, come ha proposto il Gini ⁽⁴⁾, risulta conveniente ricorrere a questa misura del potere riproduttivo che è costituita

(1) C. GINI, *Prime ricerche*, ecc., cit., pag. 334.

(2) C. GINI, *Prime ricerche*, ecc., cit., pagg. 327-329.

(3) C. GINI, op. loc. cit.

(4) Cfr.: C. GINI, *La fecondità della donna in relazione al problema della*

dal rapporto fra le nascite registrate fra il 9° ed il 12° mese dal matrimonio e il numero delle donne maritate che non hanno partorito nei primi nove mesi del matrimonio. Questo rapporto è approssimativamente uguale al rapporto delle donne che hanno concepito nei primi tre mesi del matrimonio alle donne disponibili per la proliferazione.

Questa misura del potere riproduttivo che considera anche le donne sterili, non presenta l'inconveniente dell'*I_p* quando si paragonano gruppi di donne di differenti età al matrimonio, ed inoltre presenta il vantaggio di potersi basare su un numero meno ristretto di nascite e di potere utilizzare le statistiche che forniscono distribuzioni dei primogeniti secondo la durata dal matrimonio in trimestri e non per mesi.

§ 7. — LA CONCENTRAZIONE NEL TEMPO DELLE NASCITE DEI PRIMOGENITI.

Le distribuzioni dei primogeniti classificati in funzione dell'intervallo di tempo tra il matrimonio e la nascita, possono, a nostro avviso, essere studiate anche sotto un altro aspetto. Si può infatti misurare il grado di diluizione — per così dire — delle nascite dei primogeniti nel corso del matrimonio, calcolando la *concentrazione delle nascite stesse nel tempo*, con modalità del tutto analoghe a quelle usate per la misura della concentrazione dei caratteri ⁽¹⁾.

popolazione, cit.; *Problemi della popolazione*, cit.; *Nascita Evoluzione e Morte delle Nazioni*, cit., pagg. 56 e segg.; *Le basi scientifiche della politica della popolazione*, cit., pag. 179.

(1) Cfr. C. GINI, *Variabilità e mutabilità*. Cuppini, Bologna, 1912; *Sulla misura della concentrazione e della variabilità dei caratteri*, in: «Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti», 1913-14, T. LXXIII, p. II (si veda, di queste due memorie, anche la 2ª edizione ampliata, in: *Memorie di Metodologia Statistica*, Vol. I, *Variabilità e concentrazione*, Giuffrè, Milano, 1939); A. NICEFORO, *Il Metodo Statistico*, Nuova ediz., Messina, pagg. 267-270; L. AMOROSO, *Ricerche intorno alle curve dei redditi*, in «Annali di Matematica pura ed applicata», Serie IV, vol. II, 1925; G. PIETRA, *Delle relazioni tra gl'indici di variabilità*, nota I e II, in «Atti R. Istituto Veneto S. L. A.», 1914-15, Tomo LXXIV, p. II; *Sulla teoria della variabilità nelle serie statistiche*, «Rivista Italiana di Sociologia», maggio-agosto 1915; *Nuovi contributi alla metodologia degli indici di variabilità e di concentrazione*, in «Atti del R. Istituto Veneto S. L. A.», 1931-32, Tomo XCI, p. II;

Siano N unità temporali successive $\tau_I, \tau_{II}, \dots, \tau_N$ (per esempio, i successivi giorni di un mese, i successivi mesi di calendario da una certa data in poi, ecc.), in ciascuna delle quali può verificarsi un certo evento E con la rispettiva frequenza (assoluta) f_I, f_{II}, \dots, f_N .

Ordiniamo le N unità temporali secondo l'intensità delle rispettive frequenze, facendo cioè in modo che ciascuna frequenza sia inferiore o uguale alla successiva.

Le stesse unità risulteranno perciò nel nuovo ordine :

$$\tau_1, \tau_2, \tau_3 \dots \tau_n \quad (n = N),$$

in corrispondenza alle rispettive frequenze :

$$f_1 \leq f_2 \leq f_3 \leq \dots \leq f_n.$$

Ponendo $k = 1, 2, 3, \dots, i \dots n$, si avrà evidentemente :

$$\frac{\sum_{k=1}^i f_k}{\sum_{k=1}^n f_k} \leq \frac{i}{n}; \quad [5]$$

ossia, il rapporto fra la somma delle frequenze dal posto 1 al posto i -esimo e il totale delle frequenze, è sempre inferiore o al massimo uguale al rapporto fra il numero d'ordine del posto considerato e il totale delle unità temporali. Per conseguenza, se, nella usuale rappresentazione grafica della curva di concentrazione ⁽¹⁾, si pone :

F. SAVORGNAN, *Intorno all'approssimazione di alcuni indici della distribuzione dei redditi*, in: « Atti del R. Istituto Veneto di S. L. A. », 1914-15, Tomo LXXIV, p. II; L. GALVANI, *Contributi alla determinazione degli indici di variabilità per alcuni tipi di distribuzioni*, « Metron », 1931; C. GINI, *Intorno alle curve di concentrazione*, XX^{me} Session de l'Institut International de Statistique, Madrid, 1931; L. v. BORTKIEWICZ, *Die Disparitätsmasse der Einkommens Statistik*. XIX^{me} Session de l'Institut International de Statistique. Tokio, 1930. Quest'ultima monografia rievocò alcune polemiche e discussioni intorno alle teorie degli indici di variabilità e concentrazione sorte dopo che il GINI, da oltre venticinque anni, aveva proposta la differenza media come misura della variabilità dei caratteri. In detto lavoro, apparivano dati a titolo di novità molti fra i più notevoli risultati che invece erano stati ottenuti per la prima volta dal GINI, dal PIETRA e dal SAVORGNAN. Le debite rettifiche furono inserite nel: « Bulletin de l'Institut International de Statistique », Tome XXV, Rapports et communications présentés à XIX^{me} Session, Tokio, 1930.

(1) Cfr. nota 1 a pagina precedente.

p (ascisse) = frazione del numero delle unità temporali sul totale N ($= n$) nelle quali l'ammontare del carattere (frequenze) non supera un certo limite, ossia i valori $\frac{i}{n}$ della [5];

q (ordinate) = frazione dell'ammontare totale del carattere (frequenze) che compete a quelle unità temporali, ossia i valori

$$\frac{\sum_{k=1}^i f_k}{\sum_{k=1}^n f_k}$$

della [5], si ottiene evidentemente la curva di concentrazione del fenomeno nel tempo, la quale permette di calcolare il rapporto di concentrazione con uno dei procedimenti usuali.

Nell'Appendice, alla quale rimandiamo senz'altro per maggiori dettagli, illustriamo in modo più completo questo concetto di concentrazione nel tempo di un fenomeno. Qui avvertiamo intanto che, nel caso particolare di distribuzioni *monotone decrescenti*, quali sono appunto le distribuzioni di frequenze dei primogeniti in funzione della distanza dal matrimonio, la suddetta curva di concentrazione coincide con la curva di graduazione del carattere (distanza da 9 mesi dopo il matrimonio alla nascita); ed il rapporto di concentrazione R^* che può calcolarsi in base alla curva stessa, risulta una combinazione lineare dell' I_p (cfr. Appendice).

La formola da noi adottata per il calcolo del rapporto di concentrazione nel tempo delle nascite dei primogeniti, è la seguente (cfr. Appendice, n. 7):

$$R^* = 1 - \sum_{i=0}^n (t_i + t_{i+1}) (n_{i+1} - n_i) \quad [6]$$

nella quale:

n_i = percentuale delle nascite verificatesi fino alla fine dell' i -esimo mese;

t_i = percentuale del tempo trascorso dalla 9^a ricorrenza mensile dal matrimonio fino alla fine dell' i -esimo mese, rispetto all'intervallo massimo considerato.

Alla nota (3) a pag. 94 diamo un esempio concreto del modo di impostare il calcolo di R^* per una delle nostre distribuzioni.

III.

§ 8. — IL MATERIALE UTILIZZATO.

Esaminati schematicamente i vari metodi che possono usarsi per lo studio delle distribuzioni dei primogeniti in funzione della durata del matrimonio, veniamo ad esaminare i risultati delle nostre elaborazioni. Il materiale da noi utilizzato, è per la quasi totalità, inedito, e ci è stato, su nostra richiesta, cortesemente fornito dall'Istituto Centrale di Statistica. Questo materiale ha permesso di condurre l'analisi in varie direzioni, come ad es: influenza dell'età al parto o al matrimonio sulle distribuzioni di primogeniti nei singoli Compartimenti ed in quelli a forte, media o debole natalità; ma purtroppo, moltissime altre analisi non hanno potuto essere espletate perchè i dati disponibili non risultavano rispondenti allo scopo: così ad esempio, sarebbe stato di grande interesse, ma non è stato possibile, analizzare le distribuzioni di primogeniti nelle varie categorie sociali degli sposi, nei grandi centri e negli altri comuni, ecc.

Le distribuzioni su cui si basano le nostre elaborazioni, non si riferiscono propriamente ai primogeniti nati in periodi successivi da uno stesso gruppo di matrimoni, ma sibbene ai nati in un certo periodo da matrimoni contratti in periodi anteriori diversamente remoti. L'effetto di questa circostanza sarebbe certamente trascurabile se le schiere delle donne sposatesi nei mesi successivi, fossero della stessa entità numerica e presentassero la stessa fecondabilità; ed inoltre se le donne sposate immigrate ed emigrate dal gruppo fossero anche egualmente numerose ed egualmente feconde. Tuttavia, il dubbio che siffatta circostanza possa esercitare una notevole influenza perturbatrice sui risultati delle elaborazioni, si può dire escluso dopo le ricerche eseguite dal Gini sui primogeniti nati in periodi successivi dagli stessi gruppi di matrimoni: infatti, i risultati ottenuti in base a queste distribuzioni, praticamente coincidono con quelli ottenuti in base alle distribuzioni dei primogeniti nati in un certo periodo da matrimoni diversamente remoti ⁽¹⁾.

Le nostre elaborazioni si riferiscono in parte al 1930, primo anno nel quale venne eseguita in Italia la pubblicazione dei dati

(1) C. GINI, *Nuove ricerche*, ecc., cit. pagg. 289 e 307.

relativi alle nascite dei primogeniti in funzione della distanza dal matrimonio, ed in parte al 1936, ultimo anno per il quale erano disponibili i dati stessi nel momento in cui eseguivamo i primi calcoli di questo lavoro.

Della maggioranza delle distribuzioni di primogeniti studiate, abbiamo calcolato l'intervallo protogenesico, la fecondabilità per le varie durate del matrimonio, la fecondabilità media ed il rapporto di concentrazione nel tempo, che indicheremo in seguito per brevità, rispettivamente I_p , F , R^* , e che possono assumersi come indici segnaletici delle distribuzioni medesime.

§ 9. — LE DISTRIBUZIONI DI PRIMOGENITI NEI COMPARTIMENT D'ITALIA.

Nella Tabella I riportiamo le percentuali dei primogeniti nati nel 1936 nei singoli Compartimenti, ai vari intervalli dal matrimonio, avendo escluse le nascite avvenute nei primi 9 mesi del matrimonio (concepimenti antenuziali).

TABELLA I.

Distribuzioni percentuali delle nascite dei primogeniti registrate nei Compartimenti d'Italia nel 1936 in funzione della durata del matrimonio.

COMPARTIMENTI	DURATA DEL MATRIMONIO							TOTALE
	9-12 mesi	1-2 anni	2-3 anni	3-4 anni	4-5 anni	5-9 anni	10-∞ anni	
1. Piemonte . . .	22,8	40,2	15,3	7,8	4,7	7,6	1,6	100,00
2. Liguria . . .	30,9	40,6	13,2	5,8	3,2	5,4	0,9	100,00
3. Lombardia . . .	32,1	41,7	12,4	5,1	2,8	4,9	1,0	100,00
4. Venezia Trident.	43,4	39,1	8,7	3,7	1,4	3,0	0,7	100,00
5. Veneto . . .	46,7	38,9	7,7	2,7	1,4	2,0	0,6	100,00
6. Venezia Giulia .	38,4	40,5	9,4	4,5	2,4	3,9	0,9	100,00
7. Emilia . . .	35,0	42,8	11,5	4,5	2,1	3,3	0,8	100,00
8. Toscana . . .	37,8	40,9	10,7	4,3	2,3	3,1	0,9	100,00
9. Marche . . .	45,5	40,1	7,8	2,6	1,3	2,1	0,6	100,00
10. Umbria . . .	37,1	43,8	10,7	3,2	1,7	2,7	0,8	100,00
11. Lazio . . .	45,3	39,1	7,7	3,0	1,6	2,4	0,9	100,00
12. Abruzzi . . .	40,2	41,5	9,9	3,7	1,7	2,5	0,5	100,00
13. Campania . . .	44,4	40,2	8,4	2,7	1,2	2,3	0,8	100,00
14. Puglia . . .	44,8	38,3	9,0	3,3	1,5	2,3	0,9	100,00
15. Lucania . . .	43,9	40,8	9,1	3,3	1,2	1,4	0,3	100,00
16. Calabria . . .	45,0	39,2	8,5	3,0	1,4	2,1	0,8	100,00
17. Sicilia . . .	38,5	42,7	9,6	3,7	1,7	2,9	0,9	100,00
18. Sardegna . . .	47,8	39,2	7,1	2,1	1,1	1,9	0,8	100,00

Come si vede, queste distribuzioni sono molto diverse fra loro: per citare un solo esempio, mentre nel Piemonte il 22,8 % dei primogeniti nasce nel 1° anno di matrimonio, in Sardegna, questa percentuale è più che doppia, raggiungendo il 47,8 %. È evidente poi che i Compartimenti che hanno basse percentuali di primogeniti nati a poca distanza dal matrimonio, presentano anche le più elevate percentuali di nascite a molta distanza dal matrimonio, i due fatti essendo conseguenza uno dell'altro.

Esaminiamo ora Compartimento per Compartimento gli indici segnaletici di queste distribuzioni. Nella Tabella II riportiamo per il 1930 e per il 1936 il valore dell'intervallo protogenesico $I\phi$ ⁽¹⁾, della fecondabilità media F ⁽²⁾ e del Rapporto di concentrazione nel tempo delle nascite dei primogeniti R^* ⁽³⁾.

Per il calcolo dei valori $I\phi$, F , R^* per le singole classi d'età, si sono considerate, qui come nei successivi paragrafi, solo le nascite avvenute al di sotto dei 10 anni, e ciò, sia perchè, per semplicità,

(1) Per il calcolo dell'intervallo protogenesico alle varie classi di intervallo fra matrimonio e nascita, è stato attribuito il valore centrale della classe stessa. Precisamente,

alla classe	9-10	mesi	si è attribuito il valore di mesi	9,5	-
» »	10-11	» »	» »	»	10,5
» »	11-12	» »	» »	»	11,5
» »	1-2	anni	» »	»	18
» »	2-3	» »	» »	»	30
» »	3-4	» »	» »	»	42
» »	4-5	» »	» »	»	54
» »	5-9	» »	» »	»	84
» »	10-14	» »	» »	»	144
» »	15-19	» »	» »	»	204
» »	20 ed oltre	» »	» »	»	252

Il procedimento adottato non è affatto rigoroso, in quanto ad es. le nascite avvenute fra il 1° ed il 2° anno di matrimonio non si distribuiscono certo uniformemente nel corso dell'anno ma si addensano di preferenza verso i primi mesi dell'anno. Pertanto, il considerare i valori centrali di ogni classe, porta a valutare per eccesso l'intervallo protogenesico: è evidente tuttavia che questo errore sistematico che si commette, non dovrebbe avere influenza nei confronti fra i vari compartimenti o fra i vari gruppi di donne studiati.

(2) Per il modo di calcolo della fecondabilità media, cfr. C. GINI, *Prime ricerche*, ecc., cit., pag. 334.

(3) Illustriamo con un esempio concreto il modo di calcolo del Rapporto di concentrazione nel tempo delle nascite di primogeniti. Nella se-

si può trascurare la piccolissima frazione di nati al di sopra dei 10 anni, e sia perchè, come abbiamo accennato (§ 3), con tale accorgimento si rendono abbastanza bene paragonabili per lo meno gli intervalli protogenesici relativi alle donne dai 15 ai 34 anni in quanto, sia le donne di 15, sia quelle di 34, hanno lo stesso periodo fecondo (10 anni) dinanzi ad esse per procreare il primogenito.

Esaminando la Tabella II si può osservare quanto appresso :

1) L'intervallo protogenesico I_p è molto diverso nei vari Compartimenti. Il Grafico I nel quale i Compartimenti si susseguono in ordine decrescente di intervallo per l'anno 1930, mette ancor meglio in evidenza l'andamento del fenomeno nei due anni considerati. Si può notare infatti :

guente Tabella A riportiamo le nascite di primogeniti avvenute in Piemonte nell'anno 1936 :

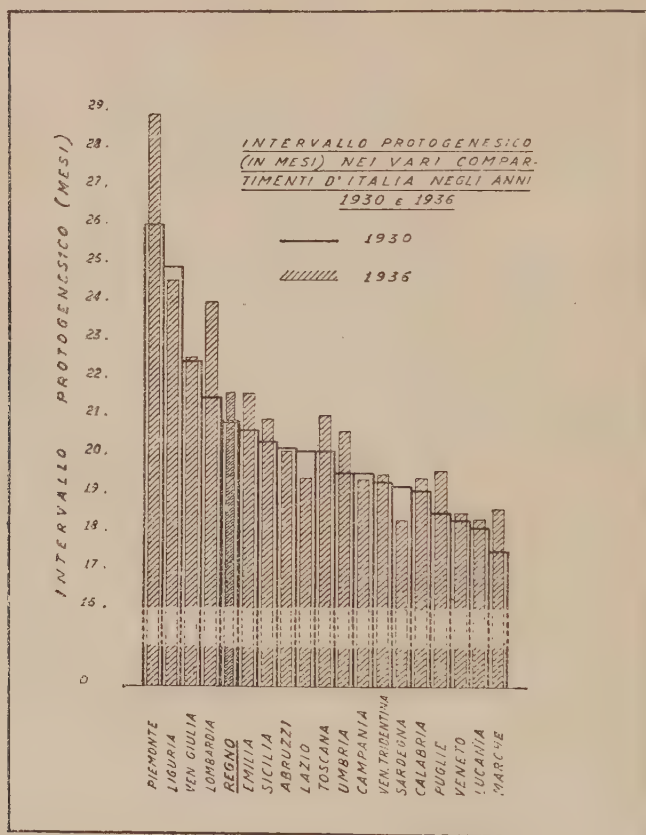
TABELLA A

MESI TRASCORSI FRA		DISTRIBUZIONE EFFETTIVA			CASO DI EQUIDISTRIBUZ.			CASO DI MASS. CONCENTRAZ.		
Matrimonio e nascita	9 ^a ricorrenza mensile dal matrimonio e nascita	nascite verificate al disotto di x mesi	100 n_i	100 t_i	nascite verificate al disotto di x mesi	100 n_i	100 t_i	nascite verificate al disotto di x mesi	100 n_i	100 t_i
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
10	I	1231	9,1	0,9	122	0,9	0,9	13596	100	0,9
11	2	2313	17,0	1,8	245	1,8	1,8	13596	100	1,8
12	3	3153	23,2	2,7	367	2,7	2,7	13596	100	2,7
24	15	8706	64,0	13,5	1835	13,5	13,5	13596	100	13,5
36	27	10817	79,6	24,3	3304	24,3	24,3	13596	100	24,3
48	39	11893	87,5	35,1	4772	35,1	35,1	13596	100	35,1
60	51	12542	92,3	46,0	6254	46,0	46,0	13596	100	46,0
120	111	13596	100,0	100,0	13596	100,0	100,0	13596	100	100,0

Esaminando la Tab. A si vede che le nascite di primogeniti avvenute al massimo entro 10 anni dal matrimonio (120 mesi) ammontano a 13.596 (col. 3). Di queste, 1231 si sono verificate al di sotto dei 10 mesi, ossia si sono verificate nel 1° mese in cui la nascita era possibile; 2313 si sono verificate nei primi 2 mesi in cui la nascita era possibile e così via, fino alle complessive 13.596 nascite di primogeniti che si sono verificate al di sotto del limite massimo di tempo considerato (120 mesi dal matrimonio; 111 mesi dal momento in cui la nascita divenne possibile = 120 — 9). Nella col. 4 sono riportati i valori 100 n_i , ossia le percentuali delle nascite, sul totale di 13.596 considerato, avvenute al di sotto di x mesi: così ad es. le 1231 nascite avvenute entro il 1° mese in cui la nascita era possibile, rappresentano il 9,1 % del totale e così via. Nella col. 5 sono indicati i valori 100 t_i , ossia le percentuali dei vari intervalli considerati: così ad es. 0,9 (1^a riag, col. 5) sta ad indicare il rapporto percentuale fra I e 111 mesi;

a) Sia nel 1930, sia nel 1936, il Piemonte, la Liguria, la Venezia Giulia e la Lombardia hanno intervallo più lungo di quello del Regno; gli altri compartimenti presentano invece intervallo più breve;

GRAFICO I



b) Dal 1930 al 1936 si ha un apprezzabile aumento di intervallo per il Regno (da mesi 20,9 a mesi 21,6);

1,8 sta ad indicare il rapporto percentuale fra 2 e 111 mesi, e così di seguito. Quindi, la coppia di valori $100 n_i = 9,1$ e $100 t_i = 0,9$, significa che in Piemonte, nell'anno 1936, il 9,1 % dei primogeniti nacque al di sotto del 0,9 % del tempo massimo considerato (1 mese su 111); la coppia di valori $100 n_i = 64,0$ e $100 t_i = 13,5$ ci dice che il 64 % impiegò per nascere un tempo inferiore al 13,5 % del tempo più lungo (15 mesi

c) *L'Ip*, dal 1930 al 1936, aumenta in Piemonte, Venezia Giulia, Lombardia, Emilia, Sicilia, Toscana, Umbria, Venezia Tridentina, Calabria, Puglia, Veneto, Lucania, Marche, mentre diminuisce lievemente in Liguria, Abruzzi, Lazio, Campania, Sardegna;

d) Per il 1930 *l'Ip* più lungo si riscontra in Piemonte (26,0), il più breve nelle Marche (17,5) con un campo di variazione di mesi 8,5. Per il 1936, *l'Ip* più lungo spetta ancora al Piemonte (28,9), il più breve alla Sardegna (18,3) con un campo di variazione di mesi 10,6.

2) Il fatto che *l'Ip* aumenta dal 1930 al 1936, in concordanza con la diminuzione della natalità generale prodottasi

su IIII, ecc. La rappresentazione grafica dei valori n_i (col. 4) e t_i (col. 5) darebbe luogo a una poligonale del tipo di quella tracciata nella Fig. 2 dell'Appendice. Nella Tabella B riportiamo il calcolo eseguito in base alla [6]

TABELLA B

MESI TRASCORSI FRA		DISTRIBUZIONE EFFETTIVA					CASO DI EQUIDISTRIBUZIONE			CASO MASSIMA CONCENTR.		
Matrimonio e nascita	9 ^a ricorrenza mensile dal matrimonio	100 n_i	100 t_i	$t_i + 1$	$n_i + 1$	$(t_i + 1) \times (n_i + 1)$	100 n_i	$n_i + 1$	$(5) \times (9)$	100 n_i	$n_i + 1$	$(5) \times (12)$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
10	I	9,1	0,9	0,9	9,1	8,2	0,9	0,9	0,8	100	0	0
11	2	17,0	1,8	2,7	7,9	21,3	1,8	0,9	2,4	100	0	0
12	3	23,2	2,7	4,5	6,2	27,9	2,7	0,9	4,1	100	0	0
24	15	64,0	13,5	16,2	40,8	661,0	13,5	10,8	175,0	100	0	0
36	27	79,6	24,3	37,8	15,6	589,7	24,3	10,8	408,2	100	0	0
48	39	87,5	35,1	59,4	7,9	469,3	35,1	10,8	641,5	100	0	0
60	51	92,3	46,0	81,1	4,8	389,3	46,0	10,9	884,0	100	0	0
120	111	100,0	100,0	146,0	7,7	1124,2	100,0	54,0	7884,0	100	0	0

per la determinazione di R^* che risulta uguale a 67,09%. Nelle coll. 8-9-10 della Tabella B è inoltre calcolata la distribuzione delle nascite (col. 8) che si otterrebbe nel caso di equidistribuzione nel tempo (Cfr. § 7 e Appendice), cioè nell'ipotesi in cui per ogni i , $n_i = t_i$. Come si vede, le nascite, anziché addensarsi verso i brevi intervalli dal matrimonio, come di fatto avviene, tenderebbero a spostarsi verso gli intervalli più lunghi.

L'ipotesi di massima concentrazione è riprodotta nelle coll. 11-12-13. In questo caso, tutte le 13.596 nascite avverrebbero al di sotto della 10^a ricorrenza mensile dal matrimonio, e la curva di concentrazione andrebbe ovviamente a coincidere coi lati del quadrato.

nello stesso periodo ⁽¹⁾, e la circostanza che i Compartimenti a più bassa natalità quali il Piemonte, la Liguria ecc., presentano più lungo intervallo, induce a pensare che fra natalità generica e ampiezza dell' $I\phi$ esiste una notevole correlazione inversa. Esamineremo in seguito (§ 10) questa correlazione più da vicino.

3) I valori $I\phi$, F , R^* , contenuti nella Tabella II, variano in notevolissima concomitanza fra loro, nel senso che i Compartimenti che hanno più lungo intervallo, presentano bassa fecondabilità media e debole concentrazione nel tempo delle nascite. Infatti il coefficiente di correlazione r per il 1930 risulta di:

— 0,908 fra l' $I\phi$ e la fecondabilità media F

— 0,995 fra l' $I\phi$ e la concentrazione R^* ⁽²⁾

+ 0,902 fra la fecondabilità media F e la concentrazione R^* .

Ciò vuol dire che in sostanza, i tre indici adoperati per lo studio delle distribuzioni dei primogeniti nei vari Compartimenti, dicono approssimativamente la stessa cosa. Ci troviamo dunque in presenza di uno di quei casi nei quali, come abbiamo accennato (§ 3), l' $I\phi$ può dare una conveniente misura della facilità al concepimento delle donne maritate che non hanno mai partorito.

Dalla Tabella II si rileva inoltre che la fecondabilità media si aggira intorno al 20 %, valore che corrisponde a quello trovato dal Gini nelle sue ricerche sulla fecondabilità della donna ⁽³⁾.

§ 10. — LE DISTRIBUZIONI DEI PRIMOGENITI IN RAPPORTO ALL'ETÀ AL PARTO.

Allo scopo di studiare il comportamento degli indici $I\phi$, F , R^* , in rapporto alla età della madre *al parto*, per il 1936, abbiamo calcolati i detti indici per i gruppi di età secondo i quali erano classificate le nascite studiate. I calcoli non sono stati eseguiti Com-

(1) Sulla diminuzione della natalità in Italia, cfr. A. NICEFORO, *Come risulta dalle Statistiche Italiane il decremento della nostra natalità*, in « Le Assicurazioni sociali », Annata XIII, luglio-agosto 1937, n. 4.

(2) Il fatto che il coefficiente di correlazione fra $I\phi$ ed R^* risulta quasi uguale all'unità, si spiega considerando che, come abbiamo avvertito (cfr. § 7 e Appendice), R^* , per le nostre distribuzioni è una combinazione lineare di $I\phi$. r non risulta esattamente eguale ad 1 per effetto degli arrotondamenti intervenuti nei calcoli.

(3) Cfr. C. GINI, *Prime ricerche*, ecc., cit., pag. 334.

TABELLA II.

Intervallo protogenesico in mesi (Ip), Fecondabilità media (F) e Concentrazione nel tempo delle nascite dei primogeniti (R) nei Compartimenti e nel Regno - 1930 e 1936 ⁽¹⁾*

	1930			1936		
	Ip	F	R*	Ip	F	R*
1. Piemonte	26,0	14,3	70,4	28,9	10,0	67,1
2. Liguria	24,8	15,8	73,5	24,6	13,6	73,6
3. Lombardia	21,5	19,2	78,3	23,9	15,3	75,2
4. Venezia Trident.	19,3	20,0	82,3	20,0	21,4	81,5
5. Veneto	18,3	26,4	83,8	18,5	22,8	83,9
6. Venezia Giulia . .	22,4	18,7	77,3	22,1	17,2	78,4
7. Emilia	20,7	20,1	79,5	21,7	16,1	78,5
8. Toscana	20,1	20,1	80,8	21,1	18,0	79,3
9. Marche	17,5	24,6	84,8	18,7	22,9	83,8
10. Umbria	19,5	20,3	82,1	20,7	17,8	80,6
11. Lazio	20,1	22,5	81,1	19,4	22,3	83,0
12. Abruzzi	20,2	20,7	81,0	20,1	19,7	81,2
13. Campania	19,5	19,7	82,1	19,4	21,1	83,2
14. Puglie	18,5	24,0	83,7	19,7	22,8	82,8
15. Lucania	18,1	23,3	84,4	18,3	22,4	84,0
16. Calabria	19,1	22,6	82,9	19,4	22,4	83,3
17. Sicilia	20,3	20,5	81,3	20,9	18,4	80,5
18. Sardegna	19,2	22,8	82,7	18,3	23,6	84,8
REGNO	20,9	—	—	21,6	—	—

partimento per Compartimento, perchè in tal caso si sarebbe prodotta un'eccessiva riduzione del numero delle osservazioni in ciascuna classe di età, che avrebbe causato forti oscillazioni accidentali specialmente sui valori *F*. Si è creduto opportuno, al contrario, studiare il fenomeno per gruppi di Compartimenti; ed al fine di poter scorgere le eventuali relazioni intercedenti fra la forma della distribuzione e la natalità generica, si sono raggruppati i 18 Compartimenti in 3 gruppi, formato ciascuno da 6 Compartimenti, dei quali il primo comprende i Compartimenti che avevano « debole » natalità nel 1936, il 2°, quelli che avevano « media » natalità, il 3°, quelli che avevano « forte » natalità ⁽²⁾.

(1) L'intervallo protogenesico, solo per questa Tabella, è stato calcolato considerando tutte le nascite, cioè anche quelle avvenute dopo il 10° anno di matrimonio.

(2) Ecco la graduatoria decrescente dei Compartimenti d'Italia per il tasso generico di natalità nel 1936:

Esaminando la seguente Tabella III nella quale sono riportati i risultati dei calcoli, ed i corrispondenti Grafici II, III e IV, si può osservare :

TABELLA III.

Intervallo protogenesico (I_p) Fecondabilità media (F) e concentrazione dei primogeniti nel tempo (R^) in gruppi di Compartimenti e per classi di età al parto (1936).*

COMPARTIMENTI CON NATALITÀ	ETÀ AL PARTO				
	15-20	21-24	25-29	30-34	35-44
<i>Intervallo protogenesico (I_p)</i>					
Debole	16,1	18,7	24,2	30,5	32,1
Media	16,2	16,9	19,4	23,2	27,2
Forte	16,1	16,8	19,1	22,2	25,6
REGNO	16,1	17,6	21,5	26,3	28,8
<i>Fecondabilità media (F)</i>					
Debole	21,0	16,4	13,4	12,5	—
Media	21,0	23,2	21,2	17,9	—
Forte	22,4	23,6	21,2	18,2	—
REGNO	21,4	20,7	17,6	15,5	—
<i>Concentrazione nel tempo (R^*)</i>					
Debole	87,2	82,4	72,1	59,7	56,6
Media	87,1	85,6	81,0	73,7	66,1
Forte	87,3	85,8	81,6	75,7	69,2
REGNO	87,2	84,4	77,1	67,7	63,0

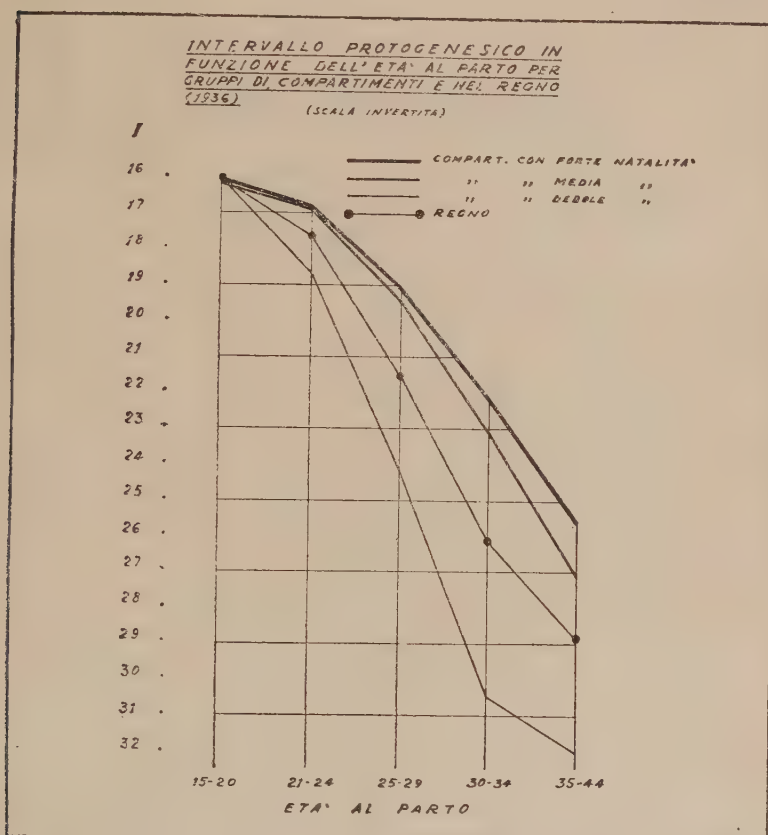
1) In tutti e tre i gruppi di Compartimenti e nel complesso del Regno, l' I_p va regolarmente e rapidamente crescendo col

Compartimenti con debole natalità	Compartimenti con media natalità	Compartimenti con forte natalità
Liguria 13,8	Ven. Trid. 20,2	Abruzzi 26,5
Piemonte 14,2	Umbria 20,8	Campania 28,3
Toscana 16,8	Marche 22,0	Sardegna 28,7
Emilia 18,3	Lazio 23,2	Puglie 30,0
Ven. Giulia 18,6	Veneto 23,3	Calabria 30,6
Lombardia 19,6	Sicilia 25,9	Lucania 32,6

Come si vede, fra il primo e l'ultimo Compartimento vi è un divario di circa 18 punti, quanti sono i Compartimenti, e ciascuno dei tre gruppi della graduatoria, abbraccia all'incirca 6 punti. È sembrato pertanto conveniente designare rispettivamente con «debole», «media» e «forte» la natalità generica dei tre gruppi suddetti.

crescere dell'età al parto (si noti che il Grafico II presenta invertita la scala delle ordinate allo scopo di poter essere facilmente comparato col Grafico IV). Andamento del tutto simile, com'era facile attendersi, hanno i valori dei Rapporti di concentrazione R^* (V. Grafico IV): infatti con l'aumentare dell'età al parto, dimi-

GRAFICO II

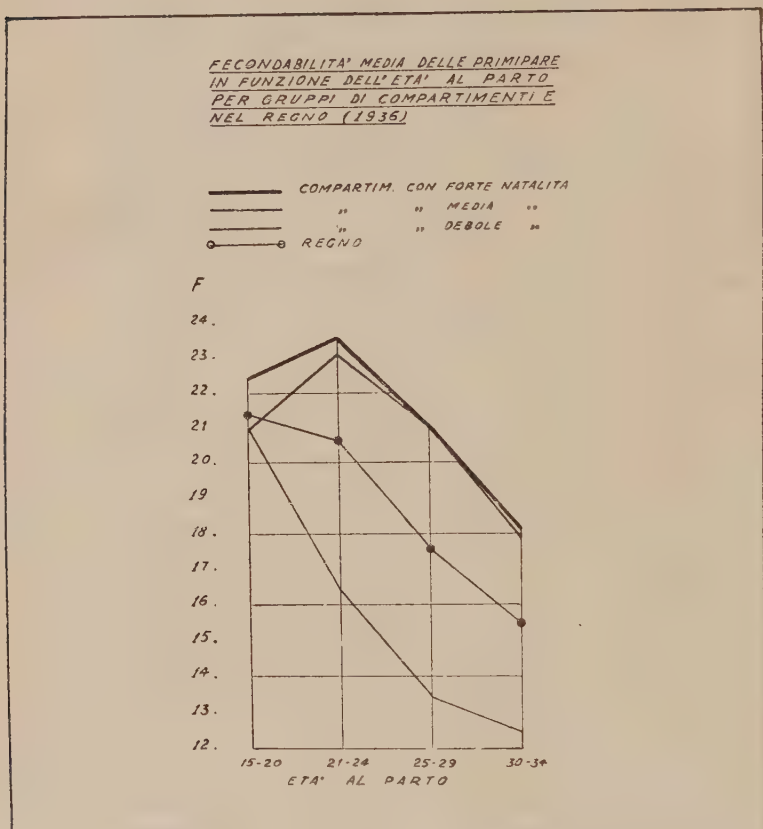


nuisce rapidamente la concentrazione, il ch  vuol dire che le donne pi  anziane al parto hanno pi  di frequente il primogenito a considerevole distanza dal matrimonio ⁽¹⁾. Naturalmente, l'aumento dell' I_p (e la conseguente diminuzione della concentrazione) col

(1) Per la classe di et  35-44 anni, non si   ritenuto opportuno calcolare il valore di F perch  il numero delle osservazioni, come abbiamo notato, per le diverse durate del matrimonio, risultava molto esiguo. Questa stessa circostanza esercita un peso meno considerevole per il calcolo di I_p e di R^* , e perci  il calcolo relativo   stato eseguito.

crescere dell'età al parto, non va senz'altro interpretata come un sintomo della diminuzione della fecondità della donna col crescere dell'età. Se si riscontra questa regolare relazione positiva tra il detto intervallo e l'età della madre, ciò è certamente, almeno in parte, perchè, a parità di condizioni, più l'intervallo è lungo,

GRAFICO III

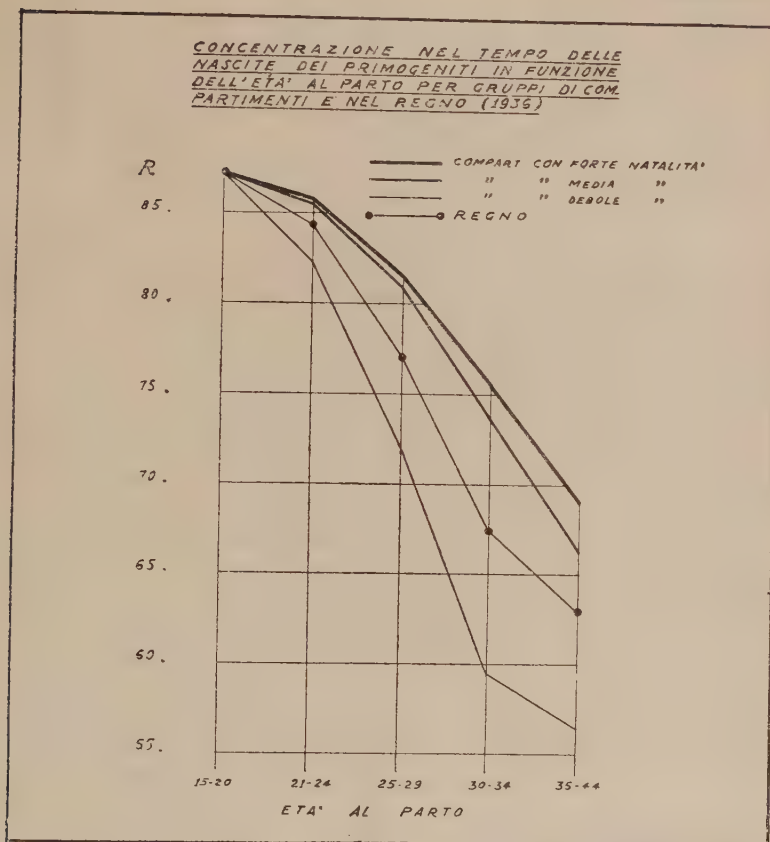


e maggiore risulta l'età della madre al parto. È anche vero, però, che le donne che partoriscono ad età più elevata, appartengono di frequente ai gruppi di donne meno fecondi: e pertanto l'aumento dell'*I_p* con l'aumento dell'età al parto, può, *fino ad un certo punto* interpretarsi come una diminuzione della fecondità della donna al crescere dell'età. Ciò viene confermato dai valori *F* che risultano in complesso anch'essi decrescenti con l'aumentare dell'età al parto (V. Graf. III). Questi risultati concordano per-

fettamente coi risultati cui giunse Knibbs ⁽¹⁾ studiando l'*I_p* in funzione dell'età al parto.

2) Per l'età 15-20, l'intervallo (e la concentrazione) risultano (V. Grafico II e IV) pressochè coincidenti nei tre gruppi di Compartimenti e nel Regno, il ch  si spiega considerando che,

GRAFICO IV



necessariamente molto uniforme deve risultare l'intervallo tra matrimonio e parto per le donne che partoriscono in giovane et . Da 21 anni in poi, invece, l'intervallo risulta sistematicamente e considerevolmente pi  lungo (e corrispondentemente la concentrazione pi  bassa) nel gruppo di Compartimenti a debole natalit , in confronto ai gruppi di Compartimenti a media e forte natalit  ;

(1) KNIBBS, op. cit., pag. 270.

ed anzi tale divario va accentuandosi con l'aumento dell'età, fino ai 30-34 anni. La spezzata relativa al Regno, ha andamento intermedio a quello delle spezzate dei tre gruppi. Tutto ciò conferma quanto avevamo accennato al § 9 circa la relazione inversa esistente fra natalità e $I\phi$ nei Compartimenti d'Italia, che del resto sarà meglio esaminata in seguito (§ 13).

3) Il Grafico III pone in evidenza le variazioni della fecondabilità media F in funzione dell'età al parto. Esso presenta una notevole particolarità rispetto ai Grafici II e IV, nel senso che la fecondabilità più alta per i Compartimenti ad alta e media natalità, si riscontra sui 21-24 anni. Ciò può considerarsi la conseguenza del fatto che, in sostanza, sotto l'età al parto, si cela, per così dire, e fino ad un certo punto, l'età al matrimonio: nel senso che le donne che hanno il primogenito ad una certa (classe di) età, appartengono in notevole misura a quella stessa classe d'età al matrimonio. Ora, come vedremo fra poco, la fecondabilità più elevata si nota, appunto nei Compartimenti ad alta e media natalità, per le donne sposate fra 21-24 anni (cfr. § 12). Rimane tuttavia da spiegare per quale ragione siffatta particolarità non appaia dai valori R^* e $I\phi$.

Per i Compartimenti a debole natalità, la fecondabilità media diminuisce rapidamente con l'aumento dell'età al parto, e pertanto il massimo di fecondabilità, anziché cadere sulla classe 21-24, come si verifica per i Compartimenti a forte e media natalità, cade sulla classe 15-20 anni. Vedremo in seguito quale interpretazione può darsi a questo fatto (cfr. §§ 12 e 13).

La spezzata relativa al Regno ha, naturalmente, andamento intermedio a quello delle spezzate relative ai tre gruppi di Compartimenti. Essa mostra diminuzione di fecondabilità dalla prima all'ultima classe considerata, il ché significa che i Compartimenti a debole natalità hanno influenza preponderante sul complesso del Regno.

Nella Tabella IV riportiamo, per le varie età al parto, l' $I\phi$ nei singoli Compartimenti. Come si vede, tanto per il 1930, quanto per il 1936, il valore di $I\phi$ nei vari Compartimenti, va crescendo col crescere dell'età al parto, il ché conferma quanto si era rilevato dalla Tabella III. Sono state in questo caso considerate tutte le età al parto, ma evidentemente, nelle classi di età avanzata, per effetto dello scarso numero delle osservazioni, le medie calcolate, hanno scarso valore rappresentativo. Inoltre, in molti Compartimenti, nelle singole classi di età, si nota un aumento del valore di $I\phi$ fra il 1930 ed il 1936.

TABELLA IV.

Intervallo protogenesico in mesi nei Compartimenti d'Italia nel 1930 e 1936

COMPARTIMENTO		A N N O 1930										A N N O 1936									
		ETÀ AL PARTO										ETÀ AL PARTO									
		TOTALE										TOTALE									
		15-20	21-24	25-29	30-34	35-39	40-44	45-50	TOTALE	15-20	21-24	25-29	30-34	35-39	40-44	45-50	TOTALE				
1	Piemonte	16,3	20,3	28,7	39,3	47,1	59,8	50,0	26,0	17,5	22,1	30,0	40,3	55,7	62,8	91,7	28,9				
2	Liguria	15,5	18,8	24,9	37,2	47,3	50,2	106,0	24,8	15,8	19,2	24,2	32,7	40,6	56,3	41,4	24,6				
3	Lombardia	14,3	16,6	22,7	34,0	46,3	60,6	41,3	21,5	11,9	17,8	23,7	34,8	49,3	60,6	40,6	23,9				
4	Venezia Tridentina	14,8	15,6	18,4	24,1	35,9	27,0	20,0	19,3	15,3	15,8	18,6	24,5	32,5	27,7	22,5	20,0				
5	Veneto	13,9	15,2	18,8	27,1	34,7	47,3	58,6	18,3	14,8	15,5	17,7	24,2	34,0	51,6	80,0	18,5				
6	Venezia Giulia e Zara	15,2	17,6	23,8	33,7	52,8	55,3	—	22,4	14,5	17,6	21,6	29,5	44,0	67,0	163,3	22,1				
7	Emilia	14,8	16,9	22,3	30,1	39,6	42,3	21,7	20,7	15,9	18,1	22,4	30,6	37,1	44,3	76,3	21,7				
8	Toscana	14,7	16,9	21,8	28,2	41,5	57,1	72,5	20,1	16,0	14,9	21,4	30,4	41,7	52,2	95,0	21,1				
9	Marche	13,9	15,6	19,5	27,4	32,8	46,9	33,3	17,5	14,9	15,7	19,8	26,1	38,9	37,9	30,0	18,7				
10	Umbria	15,2	17,3	21,0	32,0	36,6	49,4	35,0	19,5	18,1	17,8	22,3	27,3	46,3	57,7	—	20,7				
11	Lazio	15,3	14,8	22,0	29,6	41,3	39,3	80,0	20,1	15,2	16,4	20,0	26,2	39,8	49,6	31,3	19,4				
12	Abruzzi e Molise	16,8	17,9	21,8	31,6	42,9	41,7	102,5	20,2	17,4	17,6	20,7	27,4	37,3	41,0	25,0	20,1				
13	Campania	15,7	17,0	20,0	27,6	36,4	38,3	70,7	19,5	15,7	16,6	19,1	24,5	37,2	40,1	102,0	19,4				
14	Puglie	15,6	16,2	19,6	27,1	38,1	54,3	68,3	18,5	15,8	16,8	19,6	27,8	36,9	56,8	72,3	19,7				
15	Lucania	15,6	17,1	20,5	30,1	50,0	68,3	—	18,1	15,5	17,1	19,7	28,5	31,2	23,0	26,7	18,3				
16	Calabria	15,2	17,3	21,2	33,4	45,3	52,2	73,3	19,1	16,2	16,9	19,9	26,0	41,3	58,0	55,0	19,4				
17	Sicilia	16,4	18,3	22,3	29,7	42,1	63,2	108,6	20,3	16,9	18,7	22,1	29,0	36,5	51,7	266,5	20,9				
18	Sardegna	15,2	16,7	19,4	24,9	33,0	42,7	33,3	19,2	14,9	15,9	18,2	23,4	29,6	39,3	50,0	18,3				
	REGNO . . .	15,7	17,3	22,3	31,5	41,8	51,5	64,0	20,9	16,1	17,6	22,0	30,0	41,1	52,2	62,1	21,6				

N. B. — I numeri in corsivo si riferiscono a medie calcolate su un numero di casi inferiore a 50.

§ II. — VARIABILITÀ DELL'INTERVALLO PROTOGENESICO NEI COMPARTIMENTI.

Nella Tabella V riportiamo i valori della variabilità assoluta (σ) e relativa (C. V.) da noi calcolati per l'*I ϕ* ⁽¹⁾ in mesi per vari gruppi di età delle madri al parto. Come si vede, dal 1° al 2°

TABELLA V.

Variabilità assoluta (σ) e relativa (C.V.) dell'intervallo protogenesico nei Compartimenti d'Italia, per gruppi di età () al parto.*

ANNI	15-20	21-24	ETÀ AL PARTO			40-44	TOTALE
			25-29	30-34	35-39		
			Variabilità assoluta (σ)				
1930	0,77	1,29	2,41	3,93	5,75	9,99	2,16
1936	1,34	1,62	2,81	4,10	6,29	11,68	2,16
			Variabilità relativa (C. V.)				
1930	5,03	7,59	11,14	12,92	13,91	20,08	10,63
1936	8,51	9,33	13,25	14,39	15,95	23,95	12,51

(*) I valori di σ e C. V. non sono stati calcolati per la classe d'età 45-49 per il fatto che le medie riportate nella Tabella IV per questo gruppo di madri, sono state calcolate quasi tutte, su un numero molto scarso di osservazioni (cfr. annotazione alla Tab. IV).

anno, la variabilità, tanto assoluta che relativa, aumenta per il complesso delle donne e per i singoli gruppi delle varie età. Ciò potrebbe mettersi in relazione al fatto che, nello stesso periodo si è accentuata la variabilità dei Compartimenti riguardo alla natalità; difatti, il C. V. per i tassi greggi di natalità degli stessi Compartimenti risulta 21,8 % pel 1930, mentre il corrispondente valore per il 1936 raggiunge 24,0 %. Quindi, dal 1° al 2° anno si accentua l'eterogeneità dei Compartimenti, sia per la natalità che per l'ampiezza dell'*I ϕ* , che, in certo senso, ne costituisce un particolare aspetto ⁽²⁾.

(1) Naturalmente, gli indici di variabilità per i valori *F* ed *R** dovrebbero presentare un andamento analogo, dato che fra i valori *I ϕ* , *F*, *R**, esiste fortissima correlazione (cfr. § 9).

(2) Questo accentuarsi della eterogeneità dei Compartimenti attraverso il tempo per la natalità è stato del resto rilevato da tempo. Difatti, il GALVANI (L. GALVANI, *L'andamento della natalità nel settentrione e nel mezzogiorno d'Italia* ne « L'Economia Italiana », Anno XVIII, nn. 11-12, novembre-dicembre 1933, pag. 11 dell'Estratto) trovò che la natalità « mentre era verso il 1900 distribuita con una certa uniformità nelle varie regioni, ha via via dato luogo per il suo graduale diminuire in tutte, ma specialmente in alcune, a disparità di distribuzione relativamente sempre maggiori ». Difatti, i valori dei rapporti di concentrazione da questo *A* calcolati quali indici di variabilità relativa, presentano andamento nettamente crescente attraverso il tempo.

Dalla Tabella V si può rilevare inoltre che la variabilità, tanto assoluta che relativa, cresce col crescere dell'età delle madri. Ciò vuol dire, in sostanza, che i Compartimenti hanno un grado di eterogeneità maggiore quando si considera l' $I\phi$ delle donne che al parto avevano età più elevata. Questo si spiega, fino ad un certo punto, considerando che, mentre per le donne anziane al parto, a causa del loro più lungo periodo di convivenza coniugale, l'intervallo può risultare più breve o più lungo a seconda della maggiore o minore fecondità della donna — per le donne giovani al parto, al contrario, il periodo della convivenza coniugale, è necessariamente contenuto entro limiti molto più ristretti, il ch  produce appunto un restringersi del campo di possibile variazione dell' $I\phi$.

§ 12. — LE DISTRIBUZIONI DEI PRIMOGENITI IN RAPPORTO ALL'ET  AL MATRIMONIO.

Per i medesimi gruppi di Compartimenti di cui al § 10, abbiamo calcolati pel 1936 i valori $I\phi$, F , R^* , per classi classi di *et  della madre al matrimonio*. Il risultato dei calcoli,   riportato nella Tabella VI.

TABELLA VI.

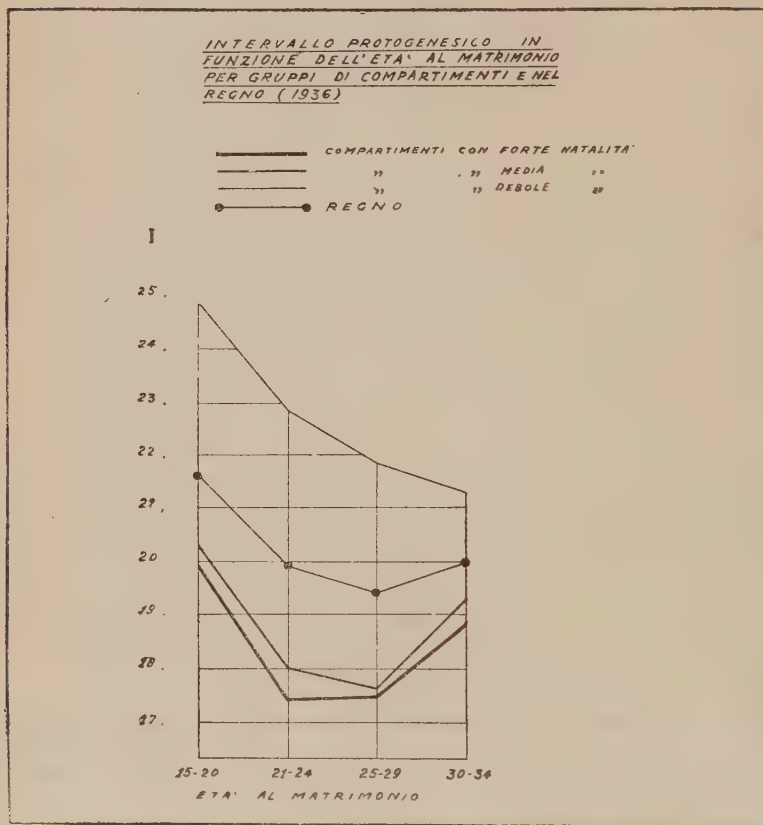
Intervallo protogenesico ($I\phi$) Fecondabilit  media (F) e Concentrazione nel tempo, (R^), in gruppi di Compartimenti e nel Regno e per classi di et  al matrimonio (1936).*

COMPARTIMENTI CON NATALIT�	ET� AL MATRIMONIO				
	15-20	21-24	25-29	30-34	35-44
<i>Intervallo Protogenesico in mesi ($I\phi$)</i>					
Debole	24,8	22,8	21,8	21,3	20,7
Media	20,3	17,9	17,6	19,3	20,4
Forte	19,9	17,4	17,5	18,8	19,9
REGNO	21,6	19,9	19,4	20,0	20,4
<i>Fecondabilit� media (F)</i>					
Debole	12,6	14,6	15,9	17,6	—
Media	16,9	22,8	23,2	20,8	—
Forte	18,6	23,6	23,3	20,6	—
REGNO	16,1	19,0	20,1	19,5	—
<i>Concentrazione nel tempo (R^*)</i>					
Debole	70,9	74,7	76,5	77,4	78,8
Media	79,3	83,9	84,1	81,2	79,2
Forte	80,0	84,6	84,5	82,1	80,1
REGNO	76,8	80,0	80,9	79,9	79,3

Esaminando i dati contenuti in questa Tabella ed i corrispondenti Grafici V, VI e VII, si può osservare :

1) Per i Compartimenti a forte e media natalità, l'intervallo minimo cade fra i 21 e i 29 anni (Grafico V). Sempre per i Compartimenti a forte e media natalità, presentano intervallo più lungo le donne che al matrimonio avevano età inferiore ai 21 anni

GRAFICO V

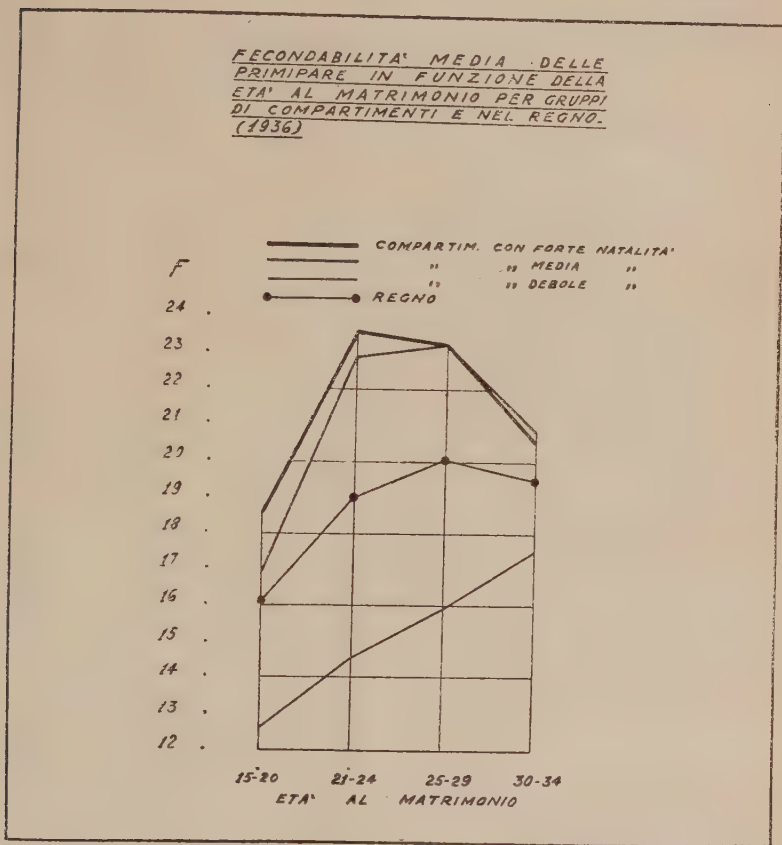


o superiore ai 29. Al contrario, nei Compartimenti a debole natalità, si verifica che l' I_p va regolarmente diminuendo con l'aumentare dell'età al matrimonio. Per le varie età, inoltre, l'intervallo è più lungo nei Compartimenti a debole natalità, intermedio, per quelli a natalità media, più piccolo per quelli a natalità forte. La spezzata relativa al Regno ha andamento intermedio alle altre

spezzate. Il Grafico VII relativo alla concentrazione, risulta quasi l'immagine rovesciata del Grafico V.

2) Il Grafico VI mostra che la fecondabilità media ha un comportamento analogo a quello dell'*I_p* e della concentrazione. Difatti: la più alta fecondabilità per i Compartimenti a media e forte natalità, si presenta fra 21 e 29 anni; per i Compartimenti

GRAFICO VI



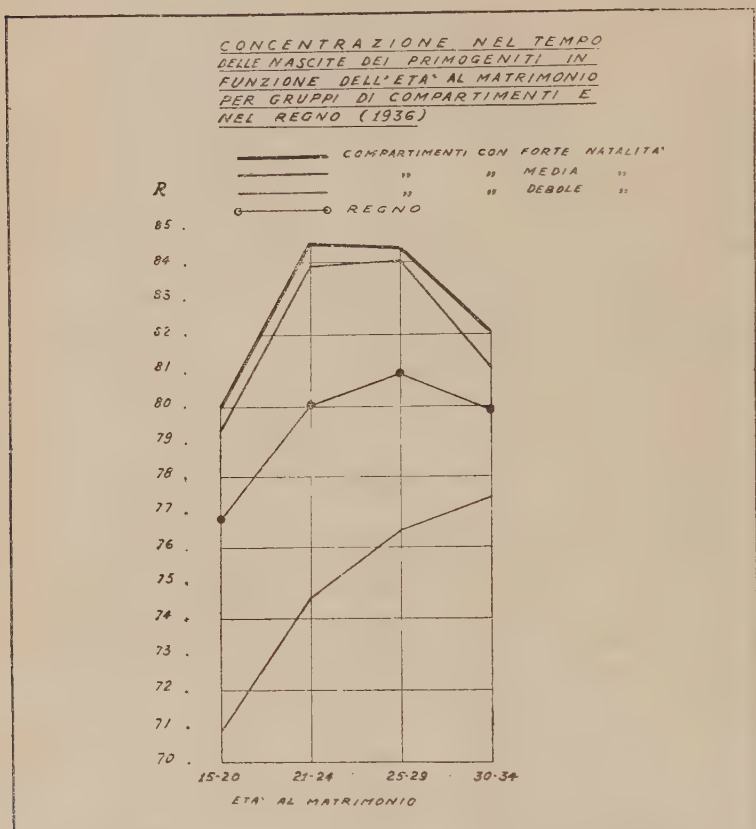
a debole natalità, la fecondabilità cresce al crescere dell'età; ed infine le donne dei Compartimenti a bassa natalità, hanno anche fecondabilità molto più bassa della corrispondente fecondabilità dei Compartimenti ad alta e media natalità generica.

3) Da quanto abbiamo detto ai precedenti nn. 1 e 2 risulta che nei compartimenti a forte natalità, la maggiore facilità

al concepimento del primogenito, si ha per le donne sposate a 21-24 anni. Tale massimo si sposta sui 25-29 anni pei compartimenti a media natalità e sulle età 30-34 pei compartimenti a debole natalità.

Se si considerano complessivamente i Compartimenti a forte e media natalità, la maggiore facilità al concepimento si ha per

GRAFICO VII



le donne sposate fra i 21 e i 29 anni. Le donne molto giovani al matrimonio (— 21 a.) o piuttosto anziane (+ 30 a.) hanno invece minore probabilità di concepire il primogenito. Questo fatto sembra concordare, grosso modo, con quanto si conosce circa le variazioni della fecondità generale della donna in rapporto all'età ⁽¹⁾.

(1) F. SAVORGNAN, *Corso di Demografia*, cit., pag. 123.

4) Il Grafico V mostra, come abbiamo detto, che pei Compartimenti a bassa natalità, l'*I ϕ* , anzichè presentare il minimo sui 21-24 o sui 25-29 anni, è regolarmente decrescente col crescere dell'età. Esso mette anche in evidenza un fatto che sembra avere una notevole importanza: il divario esistente fra l'intervallo protogenesico (e fra il rapp. di concentrazione: Graf. VII) dei Compartimenti ad alta e media natalità da una parte e l'intervallo protogenesico dei Compartimenti a debole natalità dall'altra, va rapidamente riducendosi passando dall'età minima all'età massima considerata. Le stesse caratteristiche sembrano ripetersi anche per la fecondabilità (Graf. VI), nel senso che i Compartimenti a forte e media natalità hanno il massimo di fecondabilità media sui 21-29 anni; i Compartimenti a debole natalità, hanno fecondabilità crescente col crescere dell'età; ed infine, il divario esistente tra la fecondabilità dei Compartimenti a debole natalità e quella relativa ai Compartimenti a forte e media natalità, va attenuandosi col crescere dell'età al matrimonio.

Orbene, per interpretare questo comportamento, si può fare l'ipotesi che nei Compartimenti a bassa natalità, per effetto della larga diffusione delle pratiche contraccettive, anche le spose giovani e delle età centrali (15-29 a.) mettano in atto tali pratiche per ritardare la prima nascita, sicure come sono di poter sempre avere il primogenito più tardi. A manò, a mano, invece, che l'età al matrimonio si fa più avanzata, questa sicurezza verrebbe gradatamente meno, e le pratiche contraccettive verrebbero messe più difficilmente in uso. In altre parole, nei Compartimenti a debole natalità, tenderebbero a procrastinare la prima nascita più di frequente le donne delle età giovanili e centrali al matrimonio (15-29 a.) e molto più di rado le donne anziane che temono di non poter avere più figli ritardando la prima nascita.

Il fatto poi che le donne molto giovani al matrimonio (— 21 a.) nei Compartimenti a bassa natalità abbiano l'intervallo più lungo, può interpretarsi come la conseguenza, in parte, della più bassa fecondità della donna molto giovane (cfr. sopra), e, in parte, del desiderio di ritardare la prima nascita.

L'interpretazione di cui sopra sembra la più attendibile anche perchè se il comportamento dell'intervallo (della concentrazione e della fecondabilità) pei Compartimenti a bassa natalità, si volesse imputare ad altre cause, ci sembra che difficilmente potrebbe spiegarsi, il fatto che a 30-34 anni l'*I ϕ* e la fecondabilità divengono

molto meno dissimili, rispetto alle altre età, nei vari gruppi di Compartimenti aventi diversa natalità. Per i Compartimenti a forte natalità sembra plausibile ritenere che l'intervallo rilevato, corrisponda con sufficiente approssimazione all'intervallo fisiologico, dato che in tali Compartimenti, si può pensare meno diffusa la limitazione volontaria della prole in generale ed ancora meno frequente l'abitudine di ritardare la nascita del primogenito.

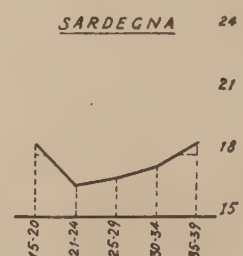
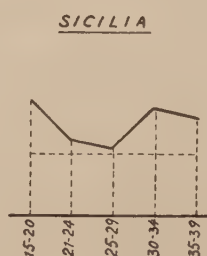
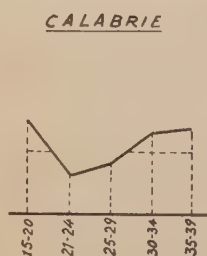
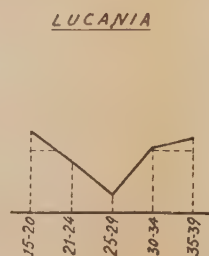
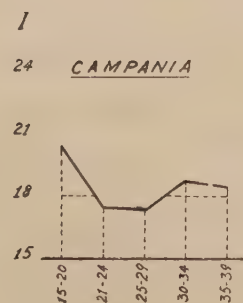
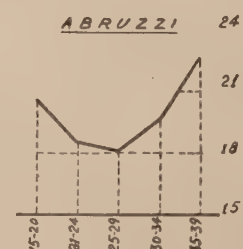
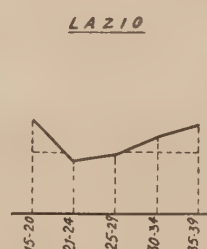
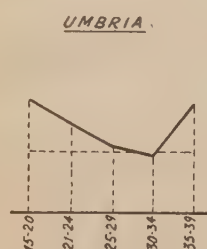
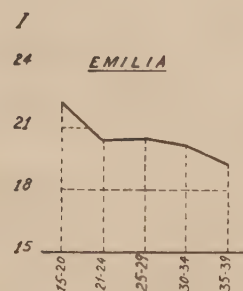
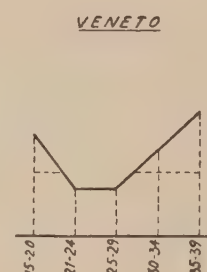
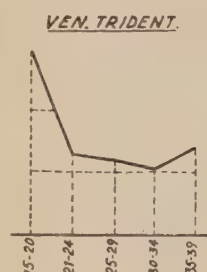
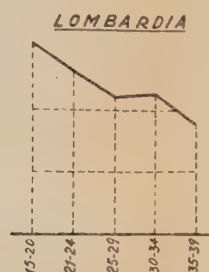
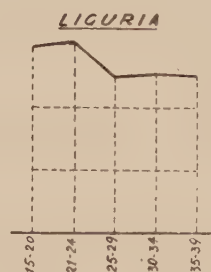
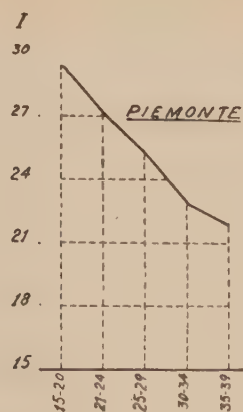
I risultati di cui si è detto, non concordano, in parte, con quelli ai quali arrivò lo Knibbs per le donne australiane nel periodo 1908-14. Egli infatti rilevò che l' I_p decresceva rapidamente dai 12 ai 19 anni, rimaneva approssimativamente costante dai 20 ai 45 anni circa, per poi diminuire rapidamente dopo tale età.

Orbene, la diminuzione dell' I_p dopo i 45 anni, si può spiegare, come del resto precisò lo stesso Knibbs (cfr. nota 1 a pag. 6) con la considerazione che le donne sposate ad età avanzata, solo alcune — le più feconde — riescono a concepire prima della menopausa, mentre le più escono dall'osservazione dopo la scomparsa delle mestruazioni. Per le età giovanili (— 21 a.) sembra esistere accordo fra i risultati cui giunse lo Knibbs ed i nostri, in quanto dalla nostra indagine appare appunto che le donne che si sposano molto giovani danno alla luce il primogenito con un certo ritardo. Per quanto concerne invece le età centrali (da 21 a 35 a.) i nostri risultati relativi al Regno non concordano con quelli dello Knibbs, perchè, come abbiamo visto, in luogo di aversi stazionarietà o quasi, il gruppo d'età 25-29 presenta I_p più breve del gruppo 30-34. Ma siffatto andamento può riconnettersi al fatto che sui nostri dati, per le donne di 25-29 anni, ha influito, molto presumibilmente, il desiderio di ritardare la nascita del primogenito mentre, è molto probabile che siffatto desiderio non abbia agito per le donne australiane del 1908-14, quando era meno diffusa la limitazione volontaria della prole.

§ 13. — L'INTERVALLO PROTOGENESICO NEI SINGOLI COMPARTIMENTI IN RELAZIONE ALL'ETÀ AL MATRIMONIO.

L'interpretazione data al § precedente alle variazioni dell' i_p (e della concentrazione e della fecondabilità) nei gruppi di Compartimenti studiati, in relazione all'età al matrimonio, richiedeva una più approfondita analisi per assumere maggior fondamento.

INTERVALLO PROTOGENESICO NEI COMPARTIMENTI D'ITALIA (1936)
IN FUNZIONE DELL'ETA' AL MATRIMONIO



A tale scopo abbiamo costruito la Tabella VII, nella quale sono riportati, Compartimento per Compartimento, gli intervalli medi in mesi fra il matrimonio e la prima nascita nei singoli gruppi di donne sposate a differenti età. Si sono considerate soltanto le donne che al matrimonio avevano al massimo 35-39 anni (cfr. § 3) e le nascite (da matrimoni puri) avvenute al di sotto di 10 anni dal matrimonio, per fare in modo che gli intervalli determinati per le varie età, fossero per quanto possibile paragonabili. Difatti per effetto di questo accorgimento, e, supponendo che la fine del periodo fecondo possa fissarsi sui 49 anni (cfr. § 3), tanto le donne

TABELLA VII.

Intervallo protogenesico nei Compartimenti d'Italia in funzione dell'età al matrimonio (nascite verificatesi al di sotto dei 10 anni dal matrimonio).

COMPARTIMENTI	ETÀ AL MATRIMONIO				
	15-20	21-24	25-29	30-34	35-39
Piemonte	29,59	27,04	25,26	22,95	21,88
Liguria	23,95	24,03	22,42	22,64	22,53
Lombardia	24,15	22,82	21,62	21,79	20,20
Venezia Tridentina .	23,98	18,89	18,55	18,13	19,02
Veneto	19,82	17,27	17,24	19,05	20,82
Venezia Giulia . . .	22,64	20,27	20,59	19,39	22,31
Emilia	22,22	20,31	20,47	20,10	19,01
Toscana	22,35	20,02	19,51	18,91	19,26
Marche	19,63	17,16	16,93	19,77	20,00
Umbria	20,61	19,33	18,35	17,87	20,20
Lazio	19,68	17,60	17,83	18,75	19,33
Abruzzi	20,62	18,50	18,03	19,54	22,63
Campania	20,38	17,41	17,32	18,72	18,36
Puglie	20,00	17,30	18,05	19,16	21,31
Lucania	18,92	17,52	15,80	18,05	18,57
Calabria	19,63	16,89	17,42	18,93	19,02
Sicilia	20,72	18,73	18,32	20,24	19,71
Sardegna	18,55	16,63	16,93	17,44	18,68

sposate a 15-20 anni, quanto quelle sposate a 35-39 anni, avevano avuto dinanzi ad esse, il medesimo periodo di convivenza coniugale fecondo (10 anni). Non è stato effettuato il calcolo per le donne di 40-44 e 45-49 anche perchè la scarshezza del numero delle osservazioni avrebbe tolto ogni valore ai risultati. Pure per la esiguità del numero delle osservazioni (in ciascun compartimento, classe di

età, e durata del matrimonio), non si è ritenuto opportuno effettuare il calcolo per le distribuzioni in questione degli indici F ed R^* ; ma l'andamento di questi — per quanto abbiamo osservato al § 9 — non dovrebbe sostanzialmente differire dall'andamento dell'Indice $I\phi$.

I dati della Tabella VII sono riportati sul Grafico VIII che mostra a colpo d'occhio quali fortissime differenze di intervallo protogenesico esistono nei vari Compartimenti e per le singole età al matrimonio. Osservando il grafico si può notare:

1) I Compartimenti a bassa natalità, come il Piemonte, la Liguria, la Lombardia, l'Emilia, la Toscana, la Venezia Giulia, hanno generalmente intervalli più lunghi dei corrispondenti intervalli degli altri Compartimenti. I medesimi Compartimenti presentano generalmente, salvo poche eccezioni di lieve entità, intervallo decrescente al crescere dell'età al matrimonio, il chè, del resto, non è altro che la conferma di quanto abbiamo rilevato al precedente paragrafo.

2) In *tutti* i Compartimenti l' $I\phi$ per l'età 15-20 è superiore a quello della età successiva (21-24) il chè induce a ritenere fondata l'osservazione fatta al paragrafo precedente (n. 3) circa la minore fecondità generale della donna molto giovane (— 21 a.). Inoltre, non è da escludere, come vedremo in seguito, (Cfr. § 16), che fra le giovanissime spose, il più frequente stato di verginità, possa produrre una minore fecondabilità della primipara (cfr. nota 2 a pag. 4).

3) L'intervallo più breve, cade nei singoli Compartimenti sulle età appresso indicate:

21-24	25-29	30-34	35-39
Lazio	Liguria	Ven. Trid.	Piemonte
Puglie	Veneto	Ven. Giulia	Lombardia
Calabria	Marche	Toscana	Emilia
Sardegna	Abruzzi	Umbria	
	Campania		
	Lucania		
	Sicilia		

Se si fa eccezione per la Liguria il cui minimo, forse a causa dello scarso numero di osservazioni, cade sui 25-29 anni, risulta che, in generale, grosso modo, l'intervallo più breve si ha per l'età 21-29 per i Compartimenti a media e forte natalità, e sulle età 30-39 anni pei Compartimenti a debole natalità.

Nel paragrafo precedente abbiamo formulato l'ipotesi che nei Compartimenti a bassa natalità tenderebbero a ritardare la prima nascita più frequentemente le donne delle età giovanili e centrali al matrimonio (15-29 anni) e molto più di rado le donne anziane al matrimonio (+ di 30 anni) che temono di non poter più avere il primogenito mettendo in atto le pratiche contraccettive. Orbene, siffatta ipotesi risulta confermata dall'esame dei coefficienti di correlazione fra natalità e $I\phi$ per le varie età al matrimonio, contenuti nella seguente Tabella VIII. Da questa

TABELLA VIII.

Coefficienti di correlazione (r) fra il tasso di natalità generica e l'intervallo protogenesico nei Compartimenti d'Italia (1936).

ETÀ AL MATRIMONIO	r	$E\text{ }p\text{ }r\text{ }(\tau)$
15-20	— 0,769	\pm 0,065
21-24	— 0,790	\pm 0,060
25-29	— 0,810	\pm 0,055
30-34	— 0,620	\pm 0,098
35-39	— 0,427	\pm 0,130

si rileva infatti che il più alto valore di r si ha per la classe 25-29 anni, mentre lo stesso valore si riduce a quasi la metà per le donne sposate a 35-39 anni. Ciò si può interpretare dicendo che le cause che tendono ad abbassare la natalità di un Compartimento, agiscono anche nel senso di allungare l'intervallo medio fra matrimonio e prima nascita: ma tale relazione inversa si affievolisce di molto per le donne sposatesi in età avanzata, il ch  significa che queste, anche se appartenenti ad un Compartimento in cui   molto diffusa la limitazione volontaria della prole, pi  difficilmente tendono a procrastinare la prima nascita.

Che le differenze nel fattore volontario abbiano una importanza notevole nel determinare le differenze da Compartimento a Compartimento dell'intervallo protogenesico alle varie et  al matrimonio, si induce anche da un'altra circostanza.

Difatti, come si rileva dalla seguente Tabella IX, la variabilit  relativa (C. V.) dell' $I\phi$ nei Compartimenti,   massima per

(1) Sullo scarso significato dell'errore probabile, cfr. C. GINI, *I pericoli della Statistica*, in: « Rivista di Politica Economica », Anno XXIX, 1939-XVIII, fasc. XI, pagg. 11 e segg. dell'Estratto.

le donne di 21-24 anni e va regolarmente diminuendo a mano a mano che aumenta l'età al matrimonio. Questo fatto, che conferma l'ipotesi di cui già si è fatto cenno (§ 12, fine) sta a significare che i Compartimenti differiscono fra loro molto meno per l' $I\phi$ delle donne anziane che per quello delle donne in età normale al matrimonio. Ora, se le differenze di intervallo nei vari Compartimenti fossero esclusivamente dovute p. es. a cause fisiologiche, non sembrerebbe doversi attendere una maggiore uniformità dei Compartimenti per riguardo all'intervallo relativo alle donne anziane al matrimonio. In altre parole, se si ammette che le differenze di $I\phi$

TABELLA IX.

Variabilità dell'intervallo protogenesico nei Compartimenti per varie età al matrimonio (1936).

ETÀ AL MATRIMONIO	σ	C. V.
15-20	2,61	12,1 %
21-24	2,71	14,1 %
25-29	2,29	12,1 %
30-34	1,51	7,8 %
35-39	1,38	6,8 %

fra Compartimento e Compartimento siano esclusivamente dovute a fattori fisiologici, non vi sono ragioni plausibili per ammettere che siffatti fattori agiscano molto più debolmente sulle donne anziane al matrimonio: chè, anzi, sarebbe forse lecito aspettarsi il contrario.

La conclusione di cui sopra non esclude però che le differenze da Compartimento a Compartimento nell' $I\phi$ siano dovute in parte anche a fattori biologici. Non esiste invero contrapposto ma accordo fra l'azione dei fattori biologici e l'azione dei fattori sociologici sulla natalità; ed anzi, il Gini ⁽¹⁾ ritiene che il fattore sociologico della razionalizzazione delle nascite, non trae esclusivamente origine da un rafforzamento *assoluto* della ragione, ma, in parte, da un rafforzamento puramente *relativo* dovuto all'indebolirsi dell'istinto. In tal modo, la indebolità

(1) C. GINI, *Le basi scientifiche della politica della popolazione*, cit., pagg. 172-177; *Nascita Evoluzione e Morte delle Nazioni*, cit., pagg. 26 e 58.

prepotenza degli istinti genetici permetterebbe alle menti di accogliere i suadenti argomenti della ragione per una regolamentazione della prole.

Vedremo più innanzi (§ 16, n. 1) come l'ipotesi che l'ampiezza dell' $I\phi$ venga determinato in parte da fattori biologici, è suggerita anche dalle differenze regionali nella fecondabilità dei primissimi mesi del matrimonio.

§ 14. — L'INTERVALLO PROTOGENESICO NEI SINGOLI COMPARTIMENTI IN RELAZIONE ALL'ETÀ AL PARTO.

Nel precedente paragrafo abbiamo visto che esiste correlazione negativa accentuata fra natalità e $I\phi$ per le età giovanili e centrali, debole per le età avanzate al matrimonio. Ora, se si considera l'età al parto anzichè al matrimonio, le cose non mutano sensibilmente, in quanto, sempre per le donne delle età centrali (al parto) si riscontrano i più alti valori di r , come si rileva dalla seguente Tabella,

TABELLA X.

Correlazione (r) fra l'intervallo protogenesico ed il tasso di natalità nei vari Compartimenti, per età al parto.

ETÀ AL PARTO	1930		1936	
	r	$E\phi r$	r	$E\phi r$
15-20	+ 0,246	\pm 0,149	+ 0,018	\pm 0,159
21-24	— 0,280	\pm 0,147	— 0,390	\pm 0,135
25-29	— 0,595	\pm 0,103	— 0,661	\pm 0,090
30-34	— 0,433	\pm 0,128	— 0,619	\pm 0,098
35-39	— 0,190	\pm 0,153	— 0,598	\pm 0,102

Anzichè considerare il tasso generico grezzo di natalità, si può assumere, quale indice del grado di fecondità del Compartimento, il numero medio di figli avuti dal complesso di tutte le donne, secondo i risultati della *Indagine sulla fecondità della donna* del 1931 ⁽¹⁾. Nella Tabella XI riportiamo appunto i valori dei coefficienti di correlazione r fra il grado di fecondità così inteso del Compartimento e l'ampiezza dell' $I\phi$. Gli elevati valori di r

(1) ISTITUTO CENTRALE DI STATISTICA DEL REGNO D'ITALIA, *Indagine sulla Fecondità della donna*. VII Censimento generale della popolazione, Vol. VI. Roma, 1936.

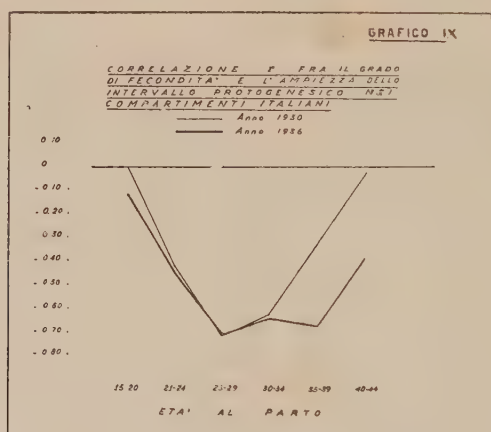
per il complesso delle donne stanno a testimoniare che, come v'era da attendersi, i Compartimenti poco fecondi hanno $I\phi$ più lungo.

TABELLA XI.

Correlazione (r) fra il grado di fecondità e l'intervallo protogenesico nei Compartimenti d'Italia.

ETÀ AL PARTO	1930			1936		
	r	$E\phi r$		r	$E\phi r$	
15-20	+ 0,008	\pm 0,159		- 0,113	\pm 0,157	
21-24	- 0,421	\pm 0,131		- 0,445	\pm 0,128	
25-29	- 0,725	\pm 0,075		- 0,711	\pm 0,079	
30-34	- 0,629	\pm 0,096		- 0,644	\pm 0,093	
35-39	- 0,332	\pm 0,141		- 0,665	\pm 0,089	
TOTALE	- 0,749	\pm 0,070		- 0,731	\pm 0,074	

Il Grafico IX mette bene in evidenza l'andamento del fenomeno per le varie età delle madri al parto, confermando, in complesso, quanto era stato rilevato dalla Tabella X. Si verifica cioè che nei



due anni considerati, la correlazione negativa più forte si ha per le donne di 25-29 anni; mentre per le donne giovani, ed anche, in parte, anziane al parto, ci si approssima quasi all'indifferenza specie se si tiene conto degli elevati errori probabili che per tali età presentano i valori di r . Ciò vorrebbe dire che per tali gruppi di donne l' $I\phi$ non risulta influenzato dalle medesime cause che producono la maggiore o minore elevatezza del grado di fecondità del Compartimento; o, in altre parole, che, variando da Compar-

mento a Compartimento questo grado di fecondità, non varia nello stesso senso o in senso opposto l'ampiezza dell'intervallo medio.

Riassumendo, l'esame della correlazione fra grado di fecondità (o tasso di natalità) del Compartimento e intervallo protogenesico per le varie età *al parto* conferma, nel complesso, le conclusioni che sono state tratte dall'esame dello stesso fenomeno per l'età al matrimonio (§ 13). Ciò si spiega considerando che, come abbiamo accennato, sotto l'età *al parto* si cela, per così dire, in molti casi, l'età al matrimonio: nel senso che le donne che hanno il primogenito ad una certa (classe di) età appartengono in notevole misura a quella stessa classe di età al matrimonio.

Una diversità notevole fra la correlazione studiata per età *al parto* e *al matrimonio*, è la seguente. Per le donne giovani al parto (15-20 a.) la correlazione è negativa e piccola in valore assoluto (Cfr. Tab. X e XI), mentre per le donne giovani al matrimonio (15-20 a.) la correlazione risulta negativa e poco più bassa di quella delle età centrali (cfr. Tab. VIII). Ciò si spiega considerando che le donne giovani al parto sono costituite da quelle schiere di donne che, sposate in giovane età, non hanno ritardato la prima nascita, per modo che l'intervallo medio ad esse relativo risulta breve ed indipendente dal grado di fecondità del Compartimento. Al contrario, le donne giovani al matrimonio dei Compartimenti nei quali è molto diffusa la limitazione volontaria, tenderebbero spesso a ritardare la prima nascita, il che spiega l'alto valore di r che per esse si desume dalla Tabella VIII.

§ 15. — LA FECONDAbilità PER LE DIVERSE DURATE DEL MATRIMONIO IN GRUPPI DI COMPARTIMENTI.

Per i medesimi gruppi di Compartimenti a forte, media e debole natalità, abbiamo calcolato per il 1936 la fecondabilità delle spose per i soliti gruppi di età al matrimonio e al parto per le varie durate del matrimonio ⁽¹⁾. Il risultato dei calcoli è riportato nella Tabella XII, cui fa riscontro il Grafico X, sul quale abbiamo riportato anche le fecondabilità medie corrispondenti già illustrate nei precedenti paragrafi.

(1) I valori delle fecondabilità per le diverse durate del matrimonio vengono riportati in questa parte della presente nota solo per comodità di esposizione. In base ad essi furono calcolati i valori della fecondabilità media, illustrati nei §§ 9 e segg. Per il metodo di calcolo dei valori p , confrontare C. GINI, *Prime ricerche*, ecc., cit., pagg. 327-329.

TABELLA XII.

Fecondabilità p () per le diverse durate del matrimonio, in gruppi di Compartimenti e per classi di età al matrimonio e al parto.*

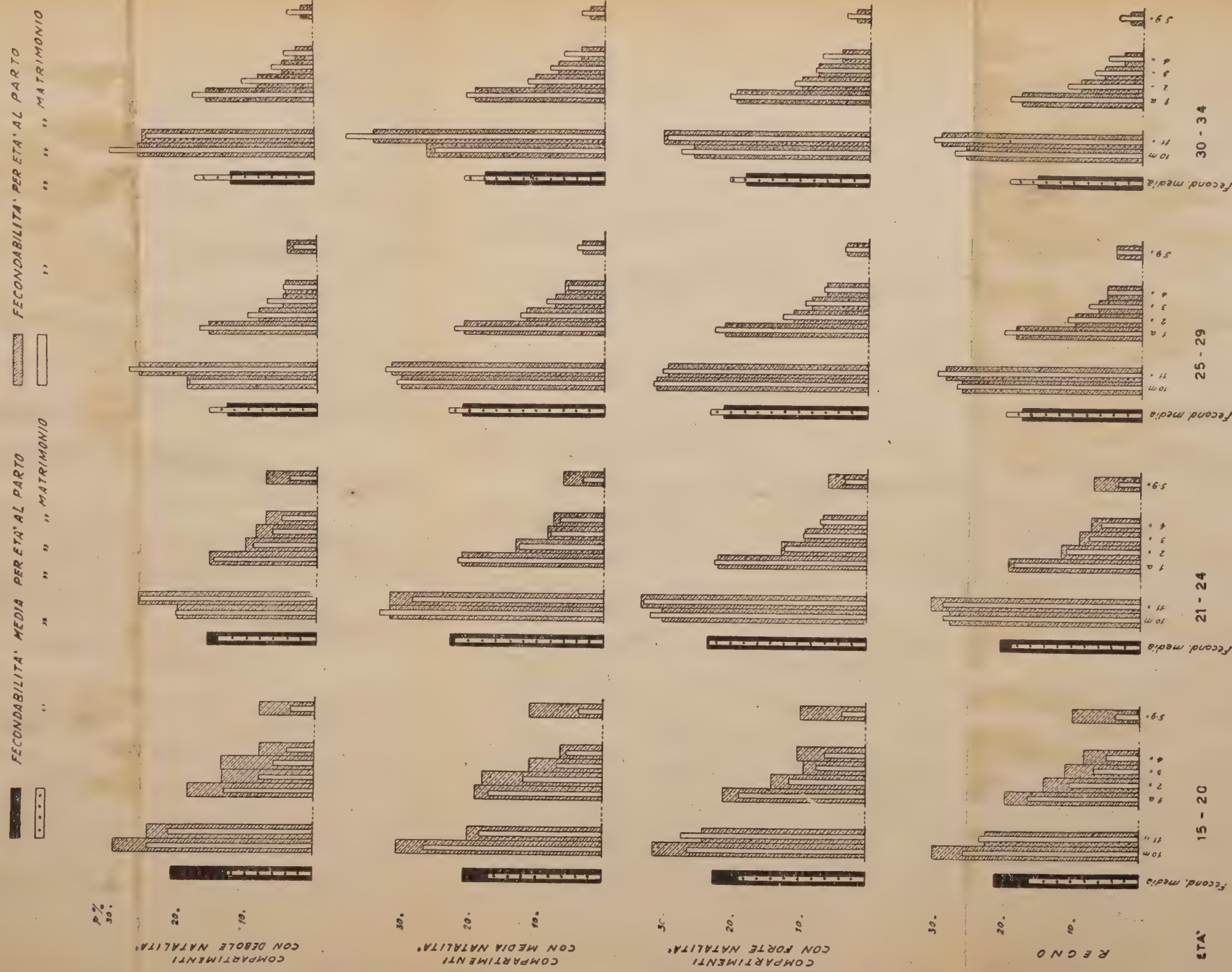
DURATA DEL MATRIMONIO	CLASSI DI ETÀ							
	AL PARTO				AL MATRIMONIO			
	15-20	21-24	25-29	30-34	15-20	21-24	25-29	30-34
<i>Compartimenti con debole natalità</i>								
10-11 mesi	29,9	20,9	19,0	26,1	24,6	20,9	19,2	30,2
11-12 mesi	24,7	26,5	26,1	25,2	21,5	26,2	27,6	24,8
1-2 anni	19,0	16,0	15,9	15,9	13,3	15,4	17,3	17,9
2-3 »	13,9	10,5	8,4	8,1	8,4	9,5	10,1	10,6
3-4 »	14,0	9,0	4,9	4,6	6,4	6,6	7,2	6,3
4-5 »	8,3	7,5	4,6	2,5	4,2	5,3	4,7	4,4
5-9 »	8,4	7,5	4,2	1,7	3,8	3,4	3,4	3,9
<i>Compartimenti con media natalità</i>								
10-11 mesi	30,9	32,3	30,6	26,6	26,7	33,6	31,1	25,4
11-12 mesi	20,3	32,2	31,9	34,5	21,1	28,9	32,8	38,4
1-2 anni	19,3	21,3	21,0	19,1	17,1	21,9	22,4	20,5
2-3 »	18,0	13,2	11,5	10,1	11,8	12,7	12,5	11,2
3-4 »	11,1	8,5	7,2	6,8	6,8	8,1	8,7	8,1
4-5 »	6,2	7,7	5,7	3,4	5,3	6,6	5,3	5,9
5-9 »	11,0	6,0	3,4	2,2	3,6	3,3	3,9	3,3
<i>Compartimenti con forte natalità</i>								
10-11 mesi	31,5	30,3	31,2	25,9	26,4	32,1	31,6	27,9
11-12 mesi	24,2	33,6	29,6	30,5	27,4	33,1	30,5	30,0
1-2 anni	21,1	22,3	21,1	19,7	18,8	22,5	22,6	20,6
2-3 »	14,0	12,7	11,1	10,1	11,2	12,3	12,5	11,2
3-4 »	9,3	9,4	8,1	7,7	7,2	9,4	9,5	8,0
4-5 »	10,3	6,6	6,3	4,2	5,9	7,0	6,4	7,0
5-9 »	9,7	5,8	3,2	2,1	3,5	3,4	3,4	3,5
<i>Regno</i>								
10-11 mesi	30,6	28,1	26,8	26,2	26,1	29,0	27,2	28,0
11-12 mesi	22,8	30,7	29,0	29,9	23,6	29,2	30,2	31,0
1-2 anni	19,9	19,5	18,5	18,0	16,5	18,5	20,4	19,5
2-3 »	14,2	11,8	9,8	9,2	10,4	11,0	11,3	11,1
3-4 »	11,0	9,0	6,4	5,8	6,7	7,5	8,0	7,2
4-5 »	8,3	7,3	5,1	3,0	4,9	5,8	5,1	5,4
5-9 »	10,0	7,0	3,8	1,9	3,6	3,4	3,5	3,7

(*) Alle diverse « durate del matrimonio » indicate in questa Tabella, corrispondono, rispettivamente le seguenti

Probabilità medie di concepimento nel corso del :

- 1° mese di matrimonio
- 2° mese di matrimonio
- 3°-15° mese di matrimonio
- 16°-28° mese di matrimonio
- 29°-41° mese di matrimonio
- 42°-54° mese di matrimonio
- 55°-115° mese di matrimonio.

FECONDABILITA' P. DELLE PRIMIPARE PER GRUPPI DI COMPARTIM. E PER CLASSI D'ETA'



Esaminando il Grafico X si può osservare :

1) La fecondabilità delle primipare (ossia la probabilità per una primipara di essere fecondata in un periodo mestruale) decresce rapidamente con l'aumentare della distanza dal matrimonio, tanto se si considera l'età al parto, quanto se si considera l'età al matrimonio.

2) La fecondabilità, a pari durata del matrimonio, se si fa eccezione specialmente per la classe 15-20 anni, risulta poco differente nelle varie classi di età e nei vari gruppi di Compartimenti, sia che si consideri l'età al parto, sia che si consideri l'età al matrimonio.

3) La fecondabilità, a pari durata del matrimonio, risulta in generale maggiore nei gruppi di Compartimenti aventi maggiore natalità.

4) La fecondabilità media per età al matrimonio in tutti e tre i gruppi di compartimenti e nel Regno, è minore della fecondabilità media per età al parto nelle classi di età 15-20 e 21-24, mentre esattamente il contrario accade per le classi di età 25-29 e 30-34. Se si esaminano poi i Grafici relativi alle età 25-29 e 30-34 per il Regno (ultimi due diagrammi in basso a destra), si rileva che alla più alta fecondabilità media per età al matrimonio, corrisponde più alta fecondabilità per età al matrimonio nelle diverse durate. Questo fatto potrebbe essere originato dalla circostanza che le donne che al parto avevano un'età compresa fra 25-34 anni, sono costituite dalla frazione meno feconda delle donne, che, sposate in precedenza, hanno il primogenito solo in quelle classi di età.

5) In molti casi si verifica che la fecondabilità del 1° mese di matrimonio, è più bassa della fecondabilità del 2° mese. Ciò può considerarsi una conferma di quanto aveva già osservato il Gini ⁽¹⁾ circa l'influenza della verginità delle spose ed altre cause che tenderebbero ad abbassare la probabilità di concepimento nel 1° mese di matrimonio (cfr. più innanzi paragrafi 16-17).

§ 16. — LA FECONDABILITÀ NEI SINGOLI COMPARTIMENTI IN RELAZIONE ALLA DURATA DEL MATRIMONIO.

Nella Tabella XIII riportiamo per il 1930 ed il 1936, i valori di p per le diverse durate del matrimonio, nei singoli Compar-

(1) C. GINI, *Prime ricerche*, ecc., cit., pag. 322.

TABELLA XIII

Fecundabilità delle primipare nei compartimenti d'Italia nel 1930 e 1936.

COMPARTIMENTI	I 9 3 0						I 9 3 6							
	Durata del matrimonio						Durata del matrimonio							
	10-11 mesi	11-12 mesi	1-2 anni	2-3 anni	3-4 anni	4-5 anni	5-9 anni	10-11 mesi	11-12 mesi	1-2 anni	2-3 anni	3-4 anni	4-5 anni	5-9 anni
Piemonte	24,3	24,8	16,4	8,2	5,2	5,3	3,0	12,1	22,4	11,3	7,7	5,5	4,1	3,6
Liguria	32,0	12,6	18,4	9,9	5,1	4,4	3,2	13,7	26,4	15,6	9,0	6,6	4,8	3,5
Lombardia	30,3	25,3	20,5	10,3	5,9	5,4	3,8	26,2	24,4	15,7	9,7	7,1	4,9	3,3
Venezia Tridentina	32,1	34,7	24,3	12,1	7,3	5,9	2,5	33,9	26,7	21,2	11,8	6,8	8,0	2,4
Veneto	37,9	36,0	25,4	12,6	7,5	4,8	3,2	30,5	31,2	22,3	12,6	8,3	5,4	4,0
Venezia Giulia	36,2	13,9	21,3	10,6	5,1	5,5	3,1	20,4	26,8	18,6	11,5	6,0	5,0	3,5
Emilia	32,7	24,9	20,8	11,0	6,9	7,1	2,3	17,6	31,8	16,6	10,4	7,6	6,1	3,7
Toscana	38,2	15,7	21,1	10,4	7,8	6,7	3,2	25,6	27,7	18,5	10,5	7,3	5,2	4,2
Marche	31,5	32,7	24,2	13,7	7,3	6,4	3,2	31,3	33,8	21,6	12,8	8,5	6,0	3,4
Umbria	35,1	14,5	21,8	11,9	6,6	6,5	3,6	33,4	19,3	17,1	11,1	9,5	5,4	3,5
Lazio	30,9	30,0	23,7	12,1	6,2	4,8	3,0	30,6	31,2	21,8	12,7	7,4	5,2	3,9
Abruzzi	31,5	29,6	20,4	11,4	6,8	6,1	3,8	27,0	32,4	19,1	11,3	7,8	6,3	3,9
Campania	26,5	24,3	20,5	11,2	8,7	6,2	3,5	27,4	27,7	21,2	12,2	9,2	6,2	3,1
Puglie	33,5	30,3	24,2	12,0	7,7	6,0	3,5	31,8	34,4	22,2	11,4	8,0	6,5	3,7
Lucania	31,7	37,9	21,5	11,6	9,9	7,7	3,9	33,4	32,5	20,8	11,8	8,1	8,4	4,5
Calabria	28,4	27,8	24,4	11,6	7,8	5,0	3,5	32,5	29,4	21,8	11,9	8,3	6,2	3,9
Sicilia	32,2	21,3	21,5	11,7	6,8	7,0	2,8	29,1	24,3	18,1	11,7	7,7	6,1	3,4
Sardegna	40,0	21,3	23,0	11,3	8,8	6,5	3,4	33,4	28,7	22,7	13,3	9,5	5,4	3,4

timenti. Il grafico XI, costruito in base ai valori della Tabella XIII, permette di osservare :

1) Esistono notevolissime differenze fra le fecondabilità relative ai vari Compartimenti, ed in generale si verifica che i più elevati valori di p si hanno per i Compartimenti ad alta e media natalità come la Venezia Tridentina, il Veneto, le Marche, gli Abruzzi, la Campania, la Puglia, la Lucania, la Calabria, la Sicilia e la Sardegna.

Questa constatazione induce a formulare il quesito se queste fortissime differenze di fecondabilità possano ricollegarsi, per lo meno in parte, a fattori di indole biologica. Ora, noi abbiamo osservato (§ 13) che vi sono fondate ragioni per ritenere che nei Compartimenti a debole natalità anche le spose novelle cercano di ritardare la prima nascita ; ma a mano a mano che l'età al matrimonio aumenta, questo desiderio si affievolisce e la fecondabilità media aumenta, tanto che le donne anziane al matrimonio, presentano fecondabilità media non molto differente a qualunque Compartimento appartengano. E poichè non sembra logico pensare che una causa fisiologica agisca per le donne giovani sì e per le anziane no, abbiamo concluso che la più bassa fecondabilità media delle spose giovani e delle età centrali dei Compartimenti a bassa natalità, debba imputarsi in parte considerevole a cause di indole sociologica. Ma, come abbiamo notato (§ 13, fine), siffatta deduzione, non porta ad escludere che queste forti differenze di fecondabilità siano dovute in parte anche a fattori biologici, i quali non debbono considerarsi in contrapposto ai fattori sociologici.

Orbene, l'esame della fecondabilità nei primi due mesi soltanto dal matrimonio ⁽¹⁾ permette di avvalorare l'ipotesi della concomitante azione delle cause biologiche. Quando infatti si paragonano i valori di p pei primi due mesi di matrimonio in due Compartimenti, il fatto che sia differente negli stessi Compartimenti la percentuale delle donne che mettono in atto pratiche contraccettive, non dovrebbe spostare i risultati che si conseguono, purchè tali percentuali rimangano costanti nel 1º, 2º e 3º mese

(1) Come abbiamo infatti ricordato (cfr. § 4), le ipotesi su cui si fonda il calcolo della fecondabilità possono ritenersi in gran parte valide quando si considerano successivi brevi intervalli ; mentre le stesse ipotesi hanno un grado di attendibilità minore per intervalli più lunghi.

di matrimonio, e le pratiche stesse raggiungano sempre il loro scopo ⁽¹⁾. Questo non è sempre vero (cfr. Gini) e perciò la fecondabilità è influenzata dalle pratiche contraccettive in quanto non abbiano pieno successo. Non si può escludere ad ogni modo che le forti differenze esistenti fra i vari Compartimenti per le fecondabilità del 1° e 2° mese di matrimonio — differenze che saranno meglio illustrate nel paragrafo successivo — possano, per lo meno in linea d'ipotesi, ricollegarsi a fattori di indole biologica ⁽²⁾.

(1) Consideriamo ad esempio le nascite di primogeniti avvenute nel 1936 nel corso del 9°, 10° e 11° mese dal matrimonio dalle donne sposate a 21-24 anni nei Compartimenti a debole e forte natalità:

	Compart. con debole natal. Nati	p	Compart. con forte natal. Nati	p
9-10 mesi.	3989		4057	
10-11 mesi.	3154	20.9	2755	32.1
11-12 mesi.	2327	26.2	1844	33.1

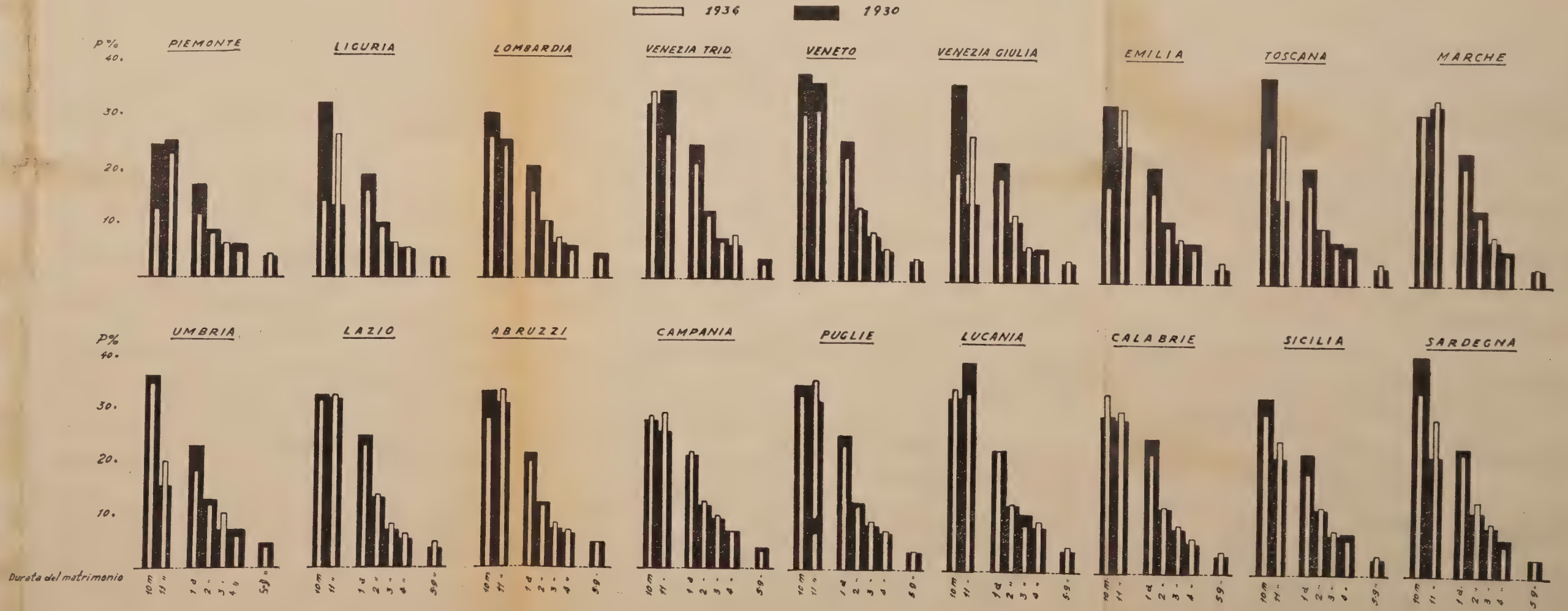
Immaginiamo che nei Compartimenti a debole natalità, nel 40% dei casi, le spose abbiano messo in atto, nel 1°, 2° e 3° mese dalle nozze delle pratiche contraccettive che abbiano impedito la fecondazione, e supponiamo che la corrispondente proporzione dei Compartimenti a forte natalità, sia del 10%. Facciamo poi l'ipotesi che le dette nascite sopra trascritte provengano esclusivamente dalle madri che non hanno opposto ostacolo alla fecondazione. Ora, è evidente che se nei due gruppi di Compartimenti quel 40% e 10% di donne non avesse del tutto cercato di limitare la prole, le nascite che si sarebbero prodotte sarebbero state:

	Compart. con debole natal. Nati	p	Compart. con forte natal. Nati	p
9-10 mesi.	5585		4463	
10-11 mesi.	4416	20.9	3031	32.1
11-12 mesi.	3258	26.2	2028	33.1

Come si vede, il valore di p rimane immutato, perchè nella [1] (§ 4) il valore di p rimane ovviamente lo stesso qualunque sia il moltiplicatore posto contemporaneamente al numeratore e al denominatore. La conseguenza è che se in due gruppi di popolazione è diversa la percentuale delle donne che nel 1°, 2°, 3° mese impediscono con risultato positivo la fecondazione, purchè siano soddisfatte le condizioni di cui sopra, i valori della fecondabilità nel 1° e nel 2° mese di matrimonio, rimangono immutati.

(2) C. GINI, *La fecondità della donna in relazione al problema della popolazione*, cit.; *Problemi della popolazione*, cit.; *Nascita, Evoluzione e Morte delle Nazioni*, cit., *Le basi scientifiche della politica della popolazione*, cit., ecc.

FECONDABILITA' DELLE PRIMIPARE NEI COMPARTIMENTI D'ITALIA - 1930 e 1936 -



Siffatta ipotesi sembrerebbe anche convalidata dalla circostanza che per i periodi molto vicini alla celebrazione del matrimonio, esiste relazione positiva relativamente bassa fra la fecondabilità nel primo o secondo mese di matrimonio e la natalità generale dei Compartimenti. Ecco i valori degli Indici di cograduazione da noi calcolati:

TABELLA XIV

Cograduazione I fra le fecondabilità per le diverse durate del matrimonio e la natalità generale (1936).

Durata del matrimonio	Cograduazione (I)
10-11 mesi.	+ 0.346
11-12 mesi.	+ 0.469
1-2 anni.	+ 0.513
2-3 anni.	+ 0.551
3-4 anni.	+ 0.581
4-5 anni.	+ 0.372
5-9 anni.	+ 0.348

Come si vede, *I* risulta massimo per la durata 3-4 anni, mentre è piuttosto basso per le durate brevi e molto lunghe ⁽¹⁾. Il basso valore di *I* per i periodi molto vicini alla celebrazione del matrimonio può interpretarsi nel senso che la elevatezza della probabilità di concepimento nei primi mesi risulta in scarsa misura connessa a quel complesso di circostanze (di natura prevalentemente sociologica) che determinano il tasso di natalità di un Compartimento. Ciò significherebbe, in altre parole, che sulle differenze di fecondabilità dei primi mesi di matrimonio nei singoli Compartimenti, hanno scarsa influenza i fattori sociologici determinanti il grado di natalità delle Regioni d'Italia; e pertanto le differenze stesse possono anche in certa misura imputarsi a fattori di indole biologica.

2) In molti Compartimenti si nota una diminuzione di fecondabilità dal 1930 al 1936 per molte durate del matrimonio.

(1) Il basso valore di *I* per le durate 4-5 e 5-9 anni, può forse attribuirsi al fatto che i valori di *p* per tali durate, differiscono di molto poco da Compartimento a Compartimento e le differenze possono anche attribuirsi in parte a cause accidentali, dato lo scarso numero di osservazioni su cui sono basati i calcoli. Non si può però ripetere la stessa osservazione per i valori che *I* assume per le durate 10-11 e 11-12 mesi.

Così ad esempio, si verifica in Piemonte, Lombardia, Venezia Tridentina, Veneto, Emilia, Toscana, Marche. In altri Compartimenti, come la Campania, si ha leggero aumento delle fecondabilità per quasi tutte le durate del matrimonio.

3) In molti Compartimenti si verifica questo fenomeno: mentre, nel 1930 la fecondabilità decresceva dal 1° al 2° mese di matrimonio, nel 1936, la fecondabilità del 2° mese, risulta più elevata di quella del 1° mese. Ciò avviene in maniera più accentuata in Liguria, Veneto, Venezia Giulia, Emilia, Toscana, Lazio, Abruzzi, Campania, Puglie.

In altri Compartimenti, la differenza fra il valore di p del 1° e 2° mese si attenua passando dal 1930 al 1936: così avviene in Lombardia, Umbria, Sardegna, Sicilia. In complesso per 14 su 18 Compartimenti, sembra avere agito una causa che avrebbe prodotto una diminuzione di fecondabilità nel 1° mese ed un aumento della fecondabilità del 2° mese. Per spiegare siffatto comportamento, si può ricorrere alla ipotesi che dal 1930 al 1936 siano andati generalizzandosi i viaggi di nozze anche per effetto delle note facilitazioni accordate dalle Ferrovie dello Stato. Il Gini, fin dal 1924 osservava infatti ⁽¹⁾: « Collegli professori di ostetricia mi informano che è particolarmente frequente, durante il 1° mese di matrimonio, il caso di mestruazioni ritardate, probabile indice di gravidanze interrotte all'inizio, fatto che essi mettono in relazione con le fatiche del viaggio di nozze e con traumi derivanti dalle coabitazioni ripetute ». Ora è evidente che se dal 1930 al 1936 si ebbe aumento del numero dei viaggi di nozze, questa circostanza potrebbe spiegare la inversione in questione.

§ 17. — LE FECONDABILITÀ NEL 1° E 2° MESE DI MATRIMONIO IN GRUPPI DI COMPARTIMENTI.

Allo scopo di esaminare quale grado di attendibilità può avere l'ipotesi della influenza dei viaggi di nozze nell'abbassare la fecondabilità nel 1° mese di matrimonio, abbiamo costruito la seguente Tabella, nella quale i Compartimenti sono raggruppati in tre Gruppi di 6 Compartimenti ciascuno, di cui il 1° comprende i Compartimenti con debole diminuzione percentuale della natalità

(1) C. GINI, *Prime ricerche ecc.*, cit., pag. 322.

dal 1930 al 1936; il 2°, i Compartimenti con media diminuzione, ed il 3° i Compartimenti con forte diminuzione ⁽¹⁾.

Dal Grafico XII nel quale sono rappresentati i valori della Tabella XV, si desume:

1) Considerando le donne di ogni età, nel 1930, la fecondabilità nel 1° mese era superiore alla fecondabilità nel 2°, ed anzi il divario andava regolarmente aumentando passando dai Compartimenti con debole a quelli con media e forte diminuzione della natalità.

2) Nel 1936, sempre per le donne di ogni età, nei Compartimenti a debole e media diminuzione della natalità, la fecondabilità nel 1° e 2° mese, tende quasi ad eguagliarsi, mentre nei Compartimenti a forte diminuzione della natalità, la fecondabilità del 2° mese, diviene decisamente superiore a quella del 1° mese, per modo che risulta invertita la situazione del 1930 ⁽²⁾. Questa stessa inversione dal 1930 al 1936 si verifica inoltre per tutte le classi di età considerate nei Compartimenti a forte diminuzione e in alcune delle classi di età dei Compartimenti a media

(1) Ecco i Compartimenti d'Italia ordinati per diminuzione percentuale della natalità dal 1930 al 1936:

Compart. con forte diminuz. natalità	Compart. con media diminuz. natalità	Compart. con debole diminuz. natalità
Umbria. 21,5 %	Marche 18,2 %	Lazio. 14,4 %
Piemonte. 21,1 %	Lombardia. . . 18,0 %	Lucania. 14,4 %
Emilia. 20,4 %	Venezia Giu. . 15,5 %	Puglie. 13,8 %
Abruzzi 19,9 %	Sicilia. 15,4 %	Campania. . . . 12,3 %
Liguria. 18,8 %	Veneto 15,3 %	Calabria 11,8 %
Toscana. 18,4 %	Venezia Trid. . 15,1 %	Sardegna. 8,6 %

(2) Calcolando il coefficiente di correlazione r fra la diminuzione percentuale della natalità dal 1930 al 1936 ed il rapporto percentuale della fecondabilità nel 1° mese (= 100) alla fecondabilità del 2° mese di matrimonio nell'anno 1936, si ottiene il valore +0,44. Questo valore sarebbe certo notevolmente più alto se non agisse la perturbazione dell'Umbria che molto presumibilmente, a causa dello scarso numero delle osservazioni, pur presentando la più forte diminuzione percentuale della natalità (21,5 %), ha rapporto molto basso (57,8). Ad ogni modo, questo valore di r sta a significare che i Compartimenti che hanno visto scendere più fortemente la loro natalità, sono quegli stessi che presentano in grado maggiore l'inversione della fecondabilità, nei primi 2 mesi del matrimonio, nel senso che in essi, nel 1936 risulta più elevata la fecondabilità nel 2° mese e più bassa quella del 1° mese, contrariamente a quanto si verificava nel 1930. Risulta pertanto confermata anche per questo verso l'interpretazione data nel testo al Grafico XII.

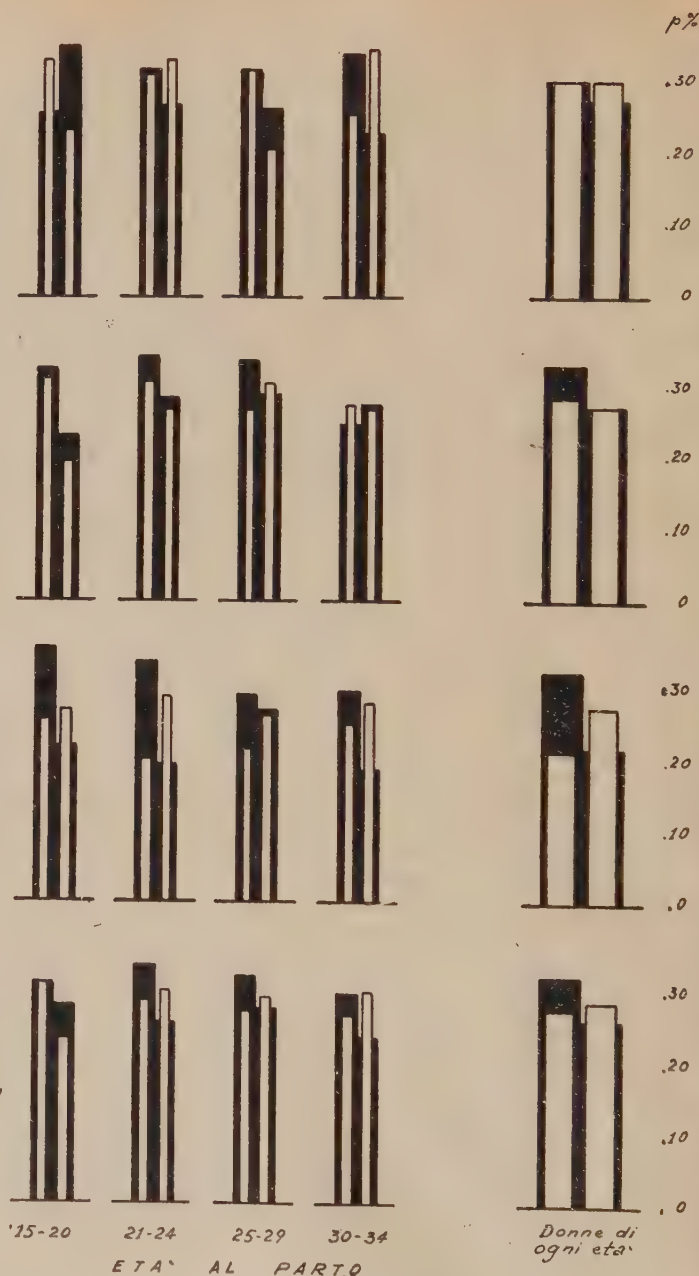
FECONDABILITA' DELLE PRIMIPARE NEL 1° E 2° MESE DI MATRIMONIO
PER GRUPPI DI COMPARTIMENTI E PER CLASSI DI ETÀ AL PARTO
1930 E 1936

COMPARTIMENTI CON
DEBOLE
 DIMINUIZIONE DELLA
 NATALITA' DAL 1930
 AL 1936

ID.
 ID. MEDIA

ID.
 ID. FORTE

REGNO
TUTTI I COMPARTIMENTI



1930
 1936

diminuzione della natalità. Laddove si verifica una diminuzione di fecondabilità nel 1° mese, sembra verificarsi quasi un aumento di compenso nel 2° mese.

3) Considerando tutti i Compartimenti insieme, si rileva ancora la stessa inversione dal 1930 al 1936, visibile non soltanto per le donne di ogni età, ma anche per le classi 21-24, 25-29, 30-34 : evidentemente, dunque, la fecondabilità dei Compartimenti con forte diminuzione della natalità, ha un peso preponderante nel determinare la fecondabilità per tutto il Regno.

TABELLA XV

Fecondabilità delle primipare nel 1° e 2° mese di matrimonio in gruppi di Compartimenti e per classi di età al parto.

FECONDABILITÀ nel	1 9 3 0					Donne di ogni età	1 9 3 6					Donne di ogni età
	Età al parto				Donne di ogni età		Età al parto				Donne di ogni età	
	15-20	21-24	25-29	30-34			15-20	21-24	25-29	30-34		
<i>Compartimenti con debole diminuzione della natalità</i>												
1° mese	25,3	31,9	31,9	34,2	30,5	33,3	31,0	31,9	25,7	30,6		
2° mese	35,0	26,8	26,3	23,1	27,9	23,2	33,1	20,7	34,6	30,4		
<i>Compartimenti con media diminuzione della natalità</i>												
1° mese	32,4	34,3	33,6	24,8	33,0	30,9	30,7	26,6	27,5	28,6		
2° mese	22,7	28,2	28,8	27,7	27,3	19,2	26,7	30,3	27,0	27,2		
<i>Compartimenti con forte diminuzione della natalità</i>												
1° mese	35,5	33,5	29,0	29,6	32,3	25,4	19,9	21,4	25,0	21,3		
2° mese	21,6	19,3	26,9	18,7	21,7	26,8	28,6	26,1	27,8	27,7		
R E G N O												
1° mese	30,6	33,3	31,8	29,3	32,0	30,6	28,1	26,8	26,2	27,5		
2° mese	27,4	25,2	27,5	23,5	25,9	22,8	29,5	29,0	29,8	28,5		

Il fatto che la diminuzione della fecondabilità nel 1° mese si manifesta in maniera evidentissima pei Compartimenti a forte diminuzione della natalità (Piemonte, Liguria, Emilia, Toscana, Umbria, Abruzzi) può spiegarsi considerando che in questi Compartimenti, per effetto di migliori condizioni economiche, abbiano avuto più larga diffusione i viaggi di nozze. Anche il fatto che l'inversione in questione non si verifica per le donne di tutti i Compartimenti dell'età di 15-20 anni, può spiegarsi con la circostanza

che le primipare di così giovane età, appartengono con maggior frequenza alle classi rurali o meno abbienti della popolazione, fra le quali il matrimonio è generalmente più precoce e meno frequentemente ci si concede il lusso del viaggio di nozze.

Noi non disponiamo di una statistica dei biglietti venduti pei viaggi di nozze nei singoli Compartimenti e negli anni 1930 e 1936. Tuttavia, ci sono stati cortesemente forniti, su nostra richiesta, dalla Direzione Generale delle Ferrovie, i seguenti dati che possono aver solo valore di larga indicazione:

TABELLA XVI

Numero dei biglietti venduti per viaggi di nozze dall'inizio della concessione (29 Luglio 1932) a tutto il 1937.

COMPARTIMENTI	Numero dei biglietti venduti	
	II Classe	III Classe
Ancona.	4.909	4.285
Bari.	3.850	1.640
Bologna.	8.965	5.493
Cagliari.	840	264
Firenze.	11.686	8.687
Genova.	7.891	2.953
Milano.	20.993	5.965
Napoli.	8.748	2.292
Palermo.	5.168	320
Reggio Calabria.	1.149	166
Torino.	10.304	4.348
Trieste.	2.600	1.052
Venezia	7.826	5.413
Roma.	15.368	6.345
REGNO . . .	110.297	49.223

Come si vede, i Compartimenti delle Ferrovie non corrispondono ai Compartimenti amministrativi, e perciò non risulta possibile calcolare dei rapporti tra il numero dei biglietti venduti ed il numero dei matrimoni celebrati nello stesso periodo pei vari Compartimenti, come sarebbe stato utile allo scopo di avere un indice della diffusione dei viaggi di nozze. Tuttavia, una semplice occhiata alle cifre su esposte, può confermare l'impressione che nelle regioni più ricche dell'Italia Settentrionale e Centrale a forte diminuzione della natalità (Piemonte, Liguria, Emilia, Toscana, Umbria, Abruzzi), più diffusi siano stati i viaggi di nozze. Orbene, questa maggiore diffusione spiegherebbe perchè, come abbiamo

accennato, l'abbassarsi della fecondabilità del 1° mese, sia più manifesto nei Compartimenti a forte diminuzione della natalità.

La presente conclusione è subordinata però alla circostanza che dal 1930 al 1936 si sia prodotto, nei vari Compartimenti, un aumento della frequenza dei viaggi di nozze. I dati della Tabella XVI non possono dare una simile risposta: tuttavia sembra logico arguire che dal 1930 (epoca nella quale la concessione non era stata ancora introdotta) al 1936, il numero dei viaggi di nozze sia andato fortemente aumentando.

§ 18. — RIASSUNTO E CONCLUSIONI.

Riassumiamo brevemente le considerazioni che precedono ed i principali risultati conseguiti:

I. — Lo studio delle distribuzioni nel tempo delle nascite dei primogeniti è di grande importanza pratica, in quanto, i risultati che si ottengono, se pure danno un'idea parziale della fecondità di una popolazione, possono ritenersi in gran parte esenti dalle manchevolezze che affettano i comuni quozienti specifici di natalità. Infatti, il gruppo delle donne studiate, risulta uniforme, sia per ciò che riguarda la probabilità di maternità derivante dal non aver mai partorito, sia per ciò che concerne l'effetto della selezione per le condizioni fisiche, ecc. Inoltre, poichè il desiderio di limitare la prole si inizia in generale dopo la nascita del primogenito, le differenze che si riscontrano nelle distribuzioni dei primogeniti in funzione dell'intervallo dal matrimonio per popolazioni differenti, possono, in certa misura, riconnettersi a fattori di indole biologica.

II. — Per studiare le distribuzioni dei primogeniti in funzione del tempo intercorso fra il matrimonio e la nascita, si possono seguire vari metodi, e cioè:

a) *Intervallo protogenesico* (I_p). — Proposto dallo Knibbs, e cioè, intervallo medio fra nascita e matrimonio. Questo indice, come ha fatto notare il Gini, può utilmente impiegarsi in certi casi (es.: I_p per spose classificate in funzione del mese di matrimonio, per gruppi di popolazione nei quali non sia apprezzabilmente differente l'età alla menopausa), mentre in altri casi, porta a risultati che non rappresentano convenientemente la facilità al concepimento (es.: I_p per spose classificate per età al matrimonio o età al parto). Con alcuni accorgimenti da noi proposti, si

può tuttavia usare anche l'intervallo protogenesico per studiare la facilità al concepimento di donne di differenti età al matrimonio.

b) *Fecondabilità (p) e fecondabilità media (F)*. — La fecondabilità, proposta dal Gini nel 1924, dà, in certe ipotesi, la probabilità che, in regime matrimoniale, all'infuori di ogni pratica contraccettiva e da ogni astensione dai contatti sessuali dirette a limitare la procreazione, la donna venga fecondata durante un periodo mestruale.

c) *Distribuzioni percentuali*. — Con questo metodo, l'esame particolareggiato delle distribuzioni può dar modo di scorgere caratteristiche che, al contrario, sfuggirebbero in base all'esame degli indici sintetici.

d) *Percentuali dei concepimenti nei primi tre mesi del matrimonio*. — Questo rapporto — proposto dal Gini — è approssimativamente uguale al rapporto fra le nascite registrate fra il 9° ed il 12° mese dal matrimonio ed il numero delle donne maritate che non hanno partorito nei primi mesi del matrimonio.

e) *Concentrazione nel tempo delle nascite dei primogeniti (R^*)*. — È questo un metodo da noi proposto e consiste essenzialmente nella estensione del concetto di concentrazione ad eventi della stessa specie che si verificano in successive unità temporali. Il valore dell'indice R^* fornisce il grado di diluizione nel tempo degli eventi considerati, risultando tanto più elevato quanto maggiore è l'addensamento degli eventi su alcune unità temporali.

Un caso particolare di concentrazione del tempo si ha per le distribuzioni *monotone decrescenti* quali sono appunto le distribuzioni delle nascite dei primogeniti in funzione della distanza dal matrimonio. Per queste distribuzioni, l'ipotesi di massima concentrazione si verificherebbe nel caso che tutte le nascite avvenissero fra il 9° e il 10° mese di matrimonio; mentre l'ipotesi di equidistribuzione si avrebbe nel caso in cui, ad es. il 25% del totale delle nascite avvenisse entro il 25% del tempo massimo considerato per la nascita del primogenito (ad es. 10 anni), il 50% delle nascite si verificasse entro il 50% del tempo massimo, ecc.

In questo caso particolare, però, la curva di concentrazione nel tempo coincide con la curva di graduazione, ed R^* risulta una combinazione lineare di I_p (cfr. Appendice).

III. — Il materiale utilizzato, quasi tutto inedito, cortesemente fornitoci dall'Istituto Centrale di Statistica, non si riferisce propriamente ai primogeniti nati in periodi successivi da uno

stesso gruppo di matrimoni, ma sibbene ai primogeniti nati in un certo periodo, da matrimoni contratti in periodi anteriori, diversamente remoti. È stato tuttavia accertato dal Gini che siffatta circostanza non sposta apprezzabilmente i risultati che si ottengono sull'uno o sull'altro materiale.

Le nostre elaborazioni si riferiscono in parte al 1930, primo anno nel quale venne eseguita in Italia l'elaborazione e la pubblicazione dei dati relativi alle nascite dei primogeniti in funzione della distanza dal matrimonio, ed in parte al 1936, ultimo anno per il quale erano disponibili i dati medesimi, nel periodo nel quale iniziavano i calcoli della presente nota. Per quasi tutte le distribuzioni disponibili, sono stati calcolati gli indici $I\phi$ (Intervallo protogenesico), F (fecondabilità media) e R^* (concentrazione nel tempo delle nascite dei primogeniti). In base ai dati disponibili, non è stato possibile calcolare la percentuale dei concepimenti nei primi tre mesi dal matrimonio, in quanto, come abbiamo ricordato, le nascite sulle quali si è operato, non provenivano dallo stesso gruppo di matrimoni, ma bensì da matrimoni diversamente remoti nel tempo.

IV. — L' $I\phi$ è molto diverso nei vari Compartimenti. Esso aumenta dal 1930 al 1936 e risulta più elevato nei Compartimenti che hanno bassa natalità. Inoltre, fra gl'indici $I\phi$, F ed R^* dei vari Compartimenti esiste fortissima correlazione, il ché significa che i tre indici, nel caso in esame, dicono tutti e tre la stessa cosa.

V. — Studiando le distribuzioni di primogeniti in funzione dell'età al parto in tre gruppi di Compartimenti (Compartimenti con debole, media, forte natalità) si osserva che col crescere dell'età al parto, aumenta l' $I\phi$ e corrispondentemente diminuisce la concentrazione, cosa che si spiega considerando che, almeno in parte, a parità di condizioni, più l'intervallo è lungo e maggiore risulta l'età al parto. È anche vero però che le donne che partoriscono in età più elevata, appartengono di frequente ai gruppi di donne meno fecondi e pertanto l'aumento dell' $I\phi$ al crescere dell'età, può, fino ad un certo punto, interpretarsi come una diminuzione della fecondità della donna al crescere dell'età. Ciò viene confermato dai valori della fecondabilità media che risultano anch'essi decrescenti con l'aumentare dell'età al parto.

VI. — Dal 1930 al 1936 si accentua l'eterogeneità dei Compartimenti Italiani tanto per la natalità generale, quanto per

I ϕ . Inoltre, nei vari Compartimenti, la variabilità di *I ϕ* cresce con l'aumentare dell'età al parto, il ch  si spiega fino ad un certo punto considerando che, mentre per le donne anziane al parto, a causa del loro pi  lungo periodo di convivenza coniugale, l'intervallo pu  risultare pi  breve o pi  lungo, per le donne giovani al parto, al contrario, il periodo della convivenza coniugale   necessariamente contenuto entro limiti molto pi  ristretti, il ch  produce appunto un restringersi del campo di variazione di *I ϕ* .

VII. — Studiando le distribuzioni dei primogeniti in funzione dell' t  al matrimonio, si osserva che l'intervallo pi  breve si ha fra i 21 e i 29 anni al matrimonio pei Compartimenti a forte e media natalit , mentre le donne pi  giovani o pi  anziane, hanno intervallo pi  lungo. Comportamento del tutto differente, hanno i Compartimenti a debole natalit : difatti, in questi, l'*I ϕ* va regolarmente decrescendo al crescere dell' t  al matrimonio. I valori *R** ed *F* confermano perfettamente questi risultati. Si deduce allora:

a) Le donne molto giovani al matrimonio (— 21 anni) o troppo anziane (+ 30 anni) nei Compartimenti a forte e media natalit  hanno pi  bassa probabilit  di concepire il primogenito.

b) Pei Compartimenti a bassa natalit , l'aumento della probabilit  di concepire il primogenito al crescere dell' t  al matrimonio, pu  interpretarsi nel senso che in questi Compartimenti, nei quali   largamente diffusa la limitazione volontaria della prole, anche le donne giovani o delle  t  centrali, tendono a ritardare la nascita del primogenito, sicure come sono di poterlo avere in seguito. Ma a mano a mano che l' t  al matrimonio si fa pi  avanzata, questa sicurezza verrebbe gradatamente meno e le pratiche contraccettive verrebbero messe pi  difficilmente in uso. Ci  risulta provato anche dai coefficienti di correlazione fra tasso generico di natalit  nei Compartimenti e intervallo protogenesico per le varie classi d' t  al matrimonio. L'influenza del fattore volontario deve essere notevole nella determinazione di questo fenomeno, perch  altrimenti non si spiegherebbe per quale ragione le donne anziane al matrimonio presentano *I ϕ* e fecondabilit  media poco differente nei tre gruppi di Compartimenti (la variabilit  dell'intervallo nei Compartimenti decresce rapidamente dall' t  21-24 all' t  35-39). Non si pu  tuttavia escludere l'ipotesi dell'influenza di fattori biologici sul fenomeno in questione (cfr. n. X).

Nei singoli Compartimenti l' *I ϕ* per l' t  15-20 al matri-

monio è sempre superiore all'intervallo dell'età successiva, il chè induce a ritenere fondata l'ipotesi sulla minore fecondità generale della donna molto giovane (— 21 anni). Ciò potrebbe essere per lo meno in parte la conseguenza del fatto che fra le giovanissime spose, una maggiore frequenza della verginità, possa produrre una minore fecondabilità delle primipare.

VIII. — Considerando l'*I ρ* nei singoli compartimenti in relazione all'età al parto ed alla natalità, si può rilevare che, nel complesso, risultano confermate le conclusioni tratte in base allo studio per l'età al matrimonio. Ciò si spiega osservando che sotto l'età al parto si cela per così dire, in molti casi, l'età al matrimonio, nel senso che le donne che hanno il primogenito ad una certa (classe di) età, appartengono in notevole misura a quella stessa classe di età al matrimonio.

IX. — La fecondabilità delle primipare decresce rapidamente con l'aumentare della distanza dal matrimonio, tanto se si considera l'età al parto, quanto se si considera l'età al matrimonio; ed inoltre, in ciascun gruppo di età e di Compartimenti (a forte, media e bassa natalità), la fecondabilità per le varie durate del matrimonio, risulta poco differente per l'età al parto ed al matrimonio.

X. — Le fecondabilità per le diverse durate del matrimonio nei singoli Compartimenti, sono molto differenti fra loro, ed in generale si verifica che i più elevati valori di *p* si hanno pei Compartimenti ad alta e media natalità. Per quanto i valori della fecondabilità media in funzione dell'età al matrimonio nei gruppi di Compartimenti inducano a pensare che sui detti valori influisca in maniera considerevole il desiderio di ritardare la nascita del primogenito (cfr. n. VII), pur tuttavia non sembra potersi escludere, per lo meno in linea d'ipotesi, che specialmente le forti differenze fra i valori della fecondabilità nel 1° e nel 2° mese di matrimonio nei vari Compartimenti, possano imputarsi a ragioni di carattere biologico. Infatti, la circostanza che in due Compartimenti sia diversa la percentuale delle spose che impediscono con effetto positivo la fecondazione, non influisce sui valori della fecondabilità del 1° e 2° mese di matrimonio, qualora siano soddisfatte determinate condizioni. D'altra parte, non esiste contrapposto fra fattori sociologici e fattori biologici, e pertanto, l'ipotesi dell'influenza di questi ultimi non si contrappone alla constatata influenza dei fattori sociologici

XI. — Analizzando più da vicino la fecondabilità del 1° e del 2° mese di matrimonio, si può osservare che, mentre nel 1930 la fecondabilità del 1° mese era superiore a quella del 2° mese, nel 1936, tale differenza si attenua, e nei Compartimenti nei quali più forte fu la discesa della natalità (dal 1930 al 1936), si produce addirittura un'inversione nel senso che risulta più elevata la fecondabilità del 2° mese, e più bassa quella del 1°. Lo stesso fenomeno si manifesta quando si considera il complesso dei Compartimenti. Questo mutamento dal 1° al 2° anno considerato della fecondabilità dei primi due mesi dal matrimonio, molto più accentuato nei Compartimenti aventi migliori condizioni economiche, sembra potersi connettere al fatto che dal 1930 al 1936 avrebbero avuto più larga diffusione i viaggi di nozze, facilitati dalla apposita concessione introdotta dal 1932 dalle Ferrovie dello Stato. Secondo opinioni di ostetrici riportate dal Gini, sembra infatti accertato che gli strapazzi causati dal viaggio di nozze, producano spesso mestruazioni ritardate, probabili indici di gravidanze interrotte all'inizio.

APPENDICE

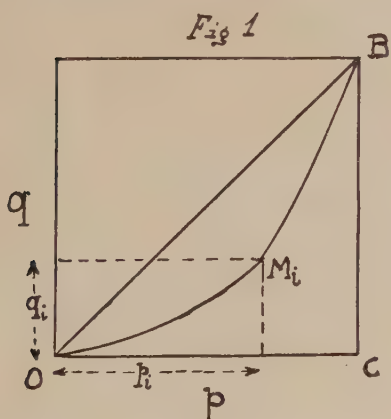
Sulla concentrazione di un fenomeno nel tempo

I. — LA CONCENTRAZIONE IN GENERALE.

È noto che la concentrazione di un carattere viene misurata mediante il *rapporto di concentrazione* del Gini ⁽¹⁾ che è dato da :

$$R = \frac{\sum_{i=1}^{n-1} (p_i - q_i)}{\sum p_i} \quad [1]$$

dove : p_i è la frazione di individui o casi in cui il carattere presenta una intensità inferiore a un dato limite ; e q_i è la frazione dell'ammontare del carattere che compete a tali casi o individui.



Se in un diagramma cartesiano ortogonale si portano sull'asse delle ascisse i valori p_i e sull'asse delle ordinate i valori q_i e si fa passare una linea continua per le estremità delle ordinate, si ottiene la ben nota *curva di concentrazione* che è ascendente da sinistra a destra e convessa rispetto all'asse delle ascisse ⁽²⁾.

Ora, si può molto facilmente dimostrare ⁽³⁾ che, quando

(1) C. GINI, *Sulla misura della concentrazione e della variabilità dei caratteri*, in: *Memorie di Metodologia statistica*, Vol. I, *Variabilità e concentrazione*, Giuffrè, Milano, 1939, pag. 364, formola 11.

(2) In luogo della curva disegnata nella fig. 1, potrebbero costruirsi altre curve, perfettamente equivalenti a quest'ultima, scambiando gli assi, ovvero considerando, anziché i valori p e q i loro complementi all'unità, cioè : $1-p$ e $1-q$ (cfr. C. GINI, *Memorie di Metodologia statistica*, cit. pag. 387).

(3) C. GINI, *Memorie di metodologia statistica*, cit. pag. 390.

cresce il numero dei casi osservati, il rapporto [1] tende al valore del rapporto dell'area compresa fra la curva di concentrazione e la retta di equidistribuzione (area di concentrazione) all'area del triangolo OBC che rappresenta l'area di concentrazione massima. (Fig. 1).

Quando — come accade quasi sempre — le intensità del carattere sono raggruppate in classi più o meno ampie, il calcolo di R viene fatto in modo approssimato con la formola ⁽¹⁾:

$$R = \frac{\sum_{k=1}^r (i_k + i_{k-1} - 1) S_k}{(n - 1) A_n} - 1 \quad [2]$$

nella quale: i_k sta ad indicare il numero degli individui o casi in cui il carattere presenta un'intensità inferiore a un dato limite; S_k , l'ammontare del carattere cadente nella classe k -esima; n , il totale degli individui o casi, ed A_n , la somma degli S_k , ossia l'ammontare totale del carattere.

Altra misura della concentrazione si ottiene calcolando l'indice di concentrazione δ in base alla formola ⁽²⁾:

$$\left[\frac{\sum_{i=n-m+1}^n a_i}{\sum_{i=1}^n a_i} \right]^{\delta} = \frac{m}{n} \quad [3]$$

dove $a_1, a_2, \dots, a_i, \dots, a_n$ sono le intensità di n fenomeni individuali dai quali risulta costituito un fenomeno collettivo A . Le intensità a_i sono graduate in modo che risulti:

$$a_i \geq a_{i-1};$$

quindi m sta a rappresentare gli ultimi m fenomeni individuali della graduatoria mentre n è il totale dei fenomeni individuali studiati e $\sum_{i=1}^n a_i$ è l'ammontare totale del carattere.

Nel caso dei redditi globali accertati e tassati il valore di δ nella [3] (formola che si adatta abbastanza bene a questi tipi di

(1) C. GINI, *Memorie di metodologia statistica*, cit., pag. 360, formola 15.

(2) C. GINI, *Memorie di metodologia statistica*, cit., pag. 12.

distribuzione), indicherà l'esponente a cui conviene innalzare la frazione dei redditi accertati e tassati posseduta dai censiti con reddito più elevato, per ottenere la frazione dei censiti che la possiede.

Il valore di δ può ottenersi applicando i logaritmi alla [3], determinando tanti valori di δ quanti sono i valori di i e facendo infine la media dei vari δ_i . Con un secondo metodo, applicando sempre i logaritmi alla [3] si perviene al valore di δ per mezzo di un'interpolazione eseguita col metodo di Cauchy ⁽¹⁾.

2. — VARIE SPECIE DI CONCENTRAZIONE: A) CONCENTRAZIONE INDIVIDUALE E FAMILIARE.

Lo stesso Gini ⁽²⁾ distingue vari tipi di indici di concentrazione in relazione ai vari modi di intendere e di studiare la concentrazione stessa. È evidente che un qualsiasi carattere quantitativo fisico, intellettuale, morale o economico può essere studiato nei riguardi della sua distribuzione fra le persone di una data popolazione, ovvero fra le famiglie di una data popolazione. Siamo nel caso di concentrazione individuale o familiare. Analogamente, saremmo nel caso di concentrazione individuale se studiassimo la concentrazione del carattere: numero dei petali in una specie di fiori, per una certa raccolta di esemplari. Nel caso di concentrazione individuale, la curva viene costruita, come abbiamo già detto (cfr. n. 1), ponendo:

p (ascisse) = frazione di individui in cui il carattere presenta un'intensità inferiore a un dato limite;
 q (ordinate) = frazione dell'ammontare del carattere che compete ai p_i individui.

3. — B) CONCENTRAZIONE TERRITORIALE.

Il Gini ⁽³⁾ propone anche di misurare la concentrazione territoriale, ossia la concentrazione di un carattere quantitativo, fra le circoscrizioni territoriali di uno Stato. Così ad esempio, se si vuole studiare il fenomeno: concentrazione della popolazione nei comuni, si pone, nella solita rappresentazione grafica:

(1) C. GINI, *Memorie di Metodologia Statistica*, cit., pag. 13-17.

(2) C. GINI, *Memorie di Metodologia Statistica*, cit. pag. 9.

(3) C. GINI, *Memorie di Metodologia Statistica*, cit. pag. 9 e 127-130.

p (ascisse) = frazione dei comuni sul totale dei comuni considerati, in cui la densità della popolazione è inferiore ad un dato limite;

q (ordinate) = frazione dell'ammontare della popolazione che compete a tali comuni.

Volendo invece applicare la [3] per ottenere l'indice di concentrazione δ , si dovrebbe porre:

n = superficie totale dello Stato;

m = superficie dei distretti a popolazione più densa di un certo limite;

$\sum_{i=1}^n a_i$ = popolazione della superficie totale;

$\sum_{i=n-m+1}^n a_i$ = popolazione della superficie dei distretti a popolazione più densa di un certo limite.

4. — C) CONCENTRAZIONE NEL TEMPO.

Il concetto di concentrazione può, a nostro avviso, essere anche esteso alle serie storiche di frequenza.

Sia $\tau_I, \tau_{II}, \dots, \tau_N$ una successione di unità temporali (per esempio: i successivi giorni di un mese; i successivi mesi di calendario da una certa data in poi, ecc.), in ciascuna delle quali può verificarsi un certo evento E con la rispettiva frequenza (assoluta) f_I, f_{II}, \dots, f_N .

Ordiniamo le N unità temporali non già secondo il loro naturale ordine di successione: I, II, III, ... N , bensì secondo l'intensità delle rispettive frequenze, facendo in modo che ogni frequenza sia inferiore o uguale alla successiva. Ogni frequenza f ed ogni unità temporale τ verranno ad assumere un nuovo posto nella nuova successione. In questa, ogni frequenza, sarà contrassegnata dall'indice I, II, III, ... N , che ci ricorda l'ordine primitivo naturale della successione; e dall'indice $1, 2, 3, \dots, n$ ($n = N$) che sta ad indicare il posto o numero d'ordine che la frequenza stessa (e la sua corrispondente unità temporale) assume nella graduatoria crescente delle frequenze. Così, date le successioni:

$$\begin{array}{c} \tau_I, \tau_{II}, \tau_{III}, \tau_{IV}, \dots, \tau_N \\ f_I, f_{II}, f_{III}, f_{IV}, \dots, f_N \end{array} \quad (a)$$

dopo aver compiuto il nuovo ordinamento, potrà risultare, ad esempio :

$$\begin{aligned} & \tau_{III,1}; \tau_{I,2}; \tau_{IV,3}; \tau_{N,4}; \dots \tau_{II,n} \\ & f_{III,1} \leq f_{I,2} \leq f_{IV,3} \leq f_{N,4} \leq \dots \leq f_{II,n} \end{aligned} \quad (b)$$

dove ad es. $\tau_{IV,3}$ significa che la 4^a unità temporale considerata nel suo naturale ordine di successione, occupa il terzo posto nella successione delle unità temporali ordinate per intensità crescenti delle rispettive frequenze; similmente $\tau_{N,4}$ significa che l'ultima unità temporale occupa il 4^o posto nella nuova successione, essendo f_N al 4^o posto nella graduatoria crescente delle frequenze, e così via.

Nelle successioni (b), considerando scritto il solo indice $1, 2, \dots, n$, si ha :

$$\begin{aligned} & \tau_1 \quad \tau_2 \quad \tau_3 \quad \dots \quad \tau_n \\ & f_1 \leq f_2 \leq f_3 \leq \dots \leq f_n \end{aligned} \quad (c)$$

Per due qualsiasi valori (i, l) di k ($k = 1, 2, \dots, i \dots n$), tali che sia $i < l$, sarà :

$$\frac{1}{i} \sum_{k=1}^i f_k \leq \frac{1}{l} \sum_{k=1}^l f_k$$

da cui :

$$\frac{\sum_{k=1}^i f_k}{\sum_{k=1}^l f_k} \leq \frac{i}{l}$$

e, nel caso particolare in cui sia $l = n$, cioè quante sono in complesso le unità temporali, si ha :

$$\frac{\sum_{k=1}^i f_k}{\sum_{k=1}^n f_k} \leq \frac{i}{n} \quad [4]$$

il ch  vuol dire che il rapporto tra la somma delle frequenze dal posto 1 al posto i -esimo e la somma totale delle frequenze,   sempre inferiore al rapporto tra il numero d'ordine del posto considerato e il totale delle unit  temporali ⁽¹⁾. Per conseguenza, se, nella consueta rappresentazione grafica, si pone :

(1) Cfr. C. GINI, *Memorie di Metodologia Statistica*, cit., pag. 361.

p (ascisse) = frazione del numero delle unità temporali sul totale N nelle quali l'ammontare del carattere (frequenze) non supera un certo limite, ossia i valori di $\frac{i}{n}$ nella [4].

q (ordinate) = frazione dell'ammontare totale del carattere (frequenze) che compete a quelle unità temporali, ossia i valori

$$\frac{\sum_{k=1}^i f_k}{\sum_{k=1}^n f_k} \text{ della [4], si ottiene evidentemente la curva di concen-}$$

trazione del fenomeno studiato nel tempo, curva che sarà sempre convessa verso l'asse delle ascisse ⁽¹⁾.

Se nella [4] si pone:

$$\sum_{k=1}^i f_k = F_i; \quad \sum_{k=1}^n f_k = F_n; \quad F_i/F_n = q_i; \quad i/n = p_i$$

si ottiene:

$$q_i \leq p_i \quad [5]$$

dove i simboli hanno il medesimo significato di quelli usati nella [1]. La concentrazione nel tempo è tanto più forte per quanto maggiore è la disuguaglianza [5], ossia, quanto maggiore è la differenza assoluta $p_i - q_i$ o la differenza relativa $\frac{p_i - q_i}{p_i}$. È per tale ragione che in generale, per avere una misura della concentrazione, si assume la [1], che non è altro che la media ponderata degli $n - 1$ valori $\frac{p_i - q_i}{p_i}$ ⁽²⁾.

La concentrazione nel tempo si dirà perfetta quando le frequenze con cui si verifica il fenomeno nelle varie unità temporali siano uguali a zero in $n - 1$ casi ed uguali ad F_n in un sol caso. Sarà allora, per gli $n - 1$ valori di i , $q_i = 0$, $p_i - q_i = p_i$, ed il valore della [1] sarà pari all'unità.

Se in tutte le unità temporali il fenomeno si presenta con la stessa frequenza, in modo che sia per tutti gli $n - 1$ valori di i , $p_i = q_i$, il numeratore della [1] sarà sempre zero, e quindi zero sarà anche il valore della intera frazione.

(1) C. GINI, *Memorie di Metodologia Statistica*, cit., pag. 387.

(2) C. GINI, *Memorie di Metodologia Statistica*, cit. pag. 364.

5. — CASO PARTICOLARE DI SERIE STORICHE DI FREQUENZA MONOTONE.

Vediamo ora che cosa avviene nel caso particolare in cui gli eventi si verificano nelle successive unità temporali con frequenze *non mai decrescenti* o *non mai crescenti*, ossia nel caso in cui le serie di frequenza studiate, siano *monotone*. Possono dunque aversi due casi :

a) Nel caso che le frequenze sono *non mai decrescenti*, si hanno le successioni :

$$\tau_{I,1}; \tau_{II,2}; \tau_{III,3}; \tau_{IV,4}; \dots; \tau_{N,k}$$

$$f_{I,1} \leq f_{II,2} \leq f_{III,3} \leq f_{IV,4} \leq \dots \leq f_{N,k}$$

cioè : la successione delle unità temporali, *rimane quella naturale* anche dopo aver compiuto l'ordinamento delle unità stesse secondo la graduatoria crescente delle rispettive frequenze.

Se, nella consueta rappresentazione grafica, si pone :

p (ascisse) = frazione del numero delle unità temporali sul totale n , nelle quali l'ammontare del carattere (frequenze) non

supera un certo limite, ossia i valori $\frac{i}{n}$ della [4] ;

q (ordinate) = frazione dell'ammontare totale del carattere (frequenze) che compete a quelle unità temporali, ossia i valori

$$\frac{\sum_{k=1}^i f_k}{\sum_{k=1}^n f_k}$$

della [4], si ottiene la curva di concentrazione del carattere (frequenze).

Il massimo di concentrazione si avrà in questo caso nell'ipotesi in cui il totale F_k delle frequenze si addensa tutto sull'ultima delle unità temporali considerate nel loro ordine naturale. L'equidistribuzione si ha nel caso in cui l'evento si verifica con la stessa frequenza nelle successive unità temporali.

b) Nel caso che le frequenze sono *non mai crescenti* (come si verifica in generale pei primogeniti quando si considerano i singoli mesi dalla 9^a ricorrenza mensile dal matrimonio in poi) si hanno le successioni :

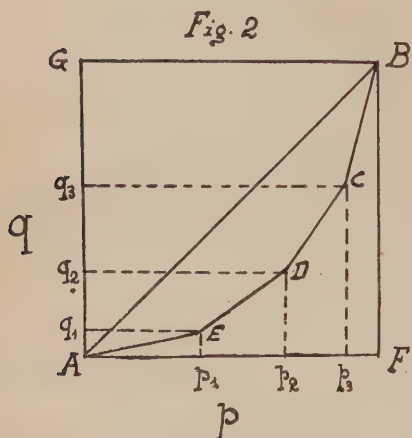
$$\tau_{N,1}; \tau_{(N-1),2}; \tau_{(N-2),3}; \dots \tau_{II,(k-1)}; \tau_{I,k}$$

$$f_{N,1} \leq f_{(N-1),2} \leq f_{(N-2),3} \leq \dots \leq f_{II,(k-1)} \leq f_{I,k}$$

cioè: la successione naturale delle unità temporali risulta *completamente invertita* dopo aver compiuto l'ordinamento delle unità stesse secondo la graduatoria crescente delle rispettive frequenze.

La curva di concentrazione sarà costruita con le stesse modalità indicate in *a*).

Il massimo di concentrazione si avrà in questo caso nell'ipotesi



in cui il totale F_k delle frequenze si addensa tutto sull'unità temporale che è prima nella successione naturale delle unità e che risulta ultima nella successione delle unità stesse dopo aver compiuto l'ordinamento in base all'intensità delle frequenze. L'equidistribuzione si avrebbe invece, come nel caso precedente, nell'ipotesi in cui l'evento si verificasse con la stessa frequenza nelle successive unità temporali.

Vedremo fra breve che — allorquando la serie storica delle

frequenze risulta non mai crescente — la curva di concentrazione delle frequenze stesse coincide con la curva di graduazione (relativa) degli eventi graduati secondo l'intervallo di tempo intercedente fra il verificarsi dell'evento considerato ed una certa origine (cfr. n. 7).

6. — IL RAPPORTO R^* DI CONCENTRAZIONE NEL TEMPO.

Una volta determinata la poligonale di concentrazione di un fenomeno nel tempo, si può misurare la concentrazione stessa nel modo seguente.

Data la poligonale iscritta nel quadrato di lato = 1 che per semplicità immaginiamo tracciata in base a soli tre punti C, D, E (fig. 2), il valore del Rapporto di concentrazione R^* ⁽¹⁾, poichè Area $ABF = 1/2$ sarà dato da :

(1) Indichiamo con R^* il rapporto di concentrazione delle frequenze nel tempo per differenziare questo indice dall'indice R del Gini, che, com'è noto, serve a misurare la variabilità relativa dei caratteri (cfr. n. 10).

$$R = \frac{1/2 - \text{Area } AEDCBF}{1/2} = 1 - 2 \text{Area } AEDCBF \quad [6]$$

Ora :

$$\text{Area } AEDCBF = \frac{p_1 q_1}{2} + \frac{q_1 + q_2}{2} (p_2 - p_1) + \frac{q_2 + q_3}{2} (p_3 - p_2) + \frac{q_3 + 1}{2} (1 - p_3)$$

dove $\frac{p_1 q_1}{2}$ può scriversi: $\frac{q_0 + q_1}{2} (p_1 - p_0)$ essendo: $q_0 = p_0 = 0$

e $\frac{q_3 + 1}{2} (1 - p_3)$ può scriversi: $\frac{q_3 + q_4}{2} (p_4 - p_3)$ essendo: $q_4 = p_4 = 1$.

Avremo dunque in generale per n punti :

$$\text{Area } AEDCBF = \frac{\sum_{i=0}^n (q_i + q_{i+1}) (p_{i+1} - p_i)}{2}$$

E, in base alla [6],

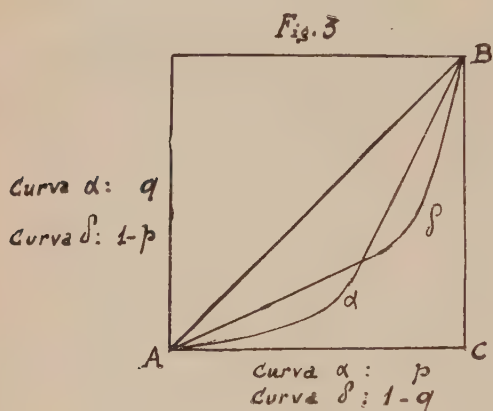
$$R = 1 - \frac{\sum_{i=0}^n (q_i + q_{i+1}) (p_{i+1} - p_i)}{1} \quad [7]$$

7. — QUANDO LA SERIE STORICA DELLE FREQUENZE RISULTA NON MAI CRESCENTE, LA CURVA DI CONCENTRAZIONE COINCIDE CON LA CURVA DI GRADUAZIONE RELATIVA DEGLI EVENTI SECONDO L'ISTANTE IN CUI ESSI SI VERIFICARONO.

Abbiamo detto che nel caso che la serie storica delle frequenze di un evento sia non mai crescente, la curva di concentrazione delle frequenze nel tempo coincide con la curva di graduazione degli eventi, costruita graduando appunto gli eventi stessi secondo l'intervallo di tempo intercedente fra l'istante in cui si verifica l'evento ed un certo istante iniziale assunto come origine comune dei tempi. Un punto generico M_i di una curva di concentrazione delle frequenze nel tempo costruita secondo le modalità illustrate (cfr. n. 5 e 6), ci dice infatti che n_i primogeniti sono stati generati al di sotto del tempo t_i dalla 9^a ricorrenza mensile dal matrimonio, — così come, un punto di una curva di graduazione per le stature dei coscritti, ci direbbe che vi è un certo numero di coscritti che ha statura inferiore ad un certo limite.

Tale coincidenza non si ha tuttavia con la curva di concentrazione α (figura 3) costruita ponendo sulle ascisse i valori di p e sulle ordinate i valori di q , ma si riscontra invece con la curva di concentrazione δ che si ottiene riportando i valori $1 - q$ sulle ascisse e i valori $1 - p$ sulle ordinate. Le due curve α e δ sono perfettamente equivalenti perchè racchiudono una medesima area di concentrazione ⁽¹⁾.

Nella Tabella A è riportato un esempio numerico che mostra appunto la perfetta corrispondenza della curva di concentrazione



delle frequenze δ con la curva di graduazione relativa degli eventi (nascite dei primogeniti) secondo l'istante in cui essi si verificano. Tale curva si ottiene ponendo uguale all'unità sia il totale delle frequenze (ascisse), sia il valore più elevato del carattere (intervallo di tempo fra la 9^a ricorrenza mensile dal matrimonio e la nascita del primogenito: ordinate).

L'esempio si riferisce alle nascite dei primogeniti avvenute entro 5 anni dal matrimonio, in Sardegna nel 1936: il numero delle nascite nei singoli mesi è stato ottenuto dividendo le frequenze d'ogni classe per il numero dei mesi di cui era composta la classe.

Come si vede dalla Tabella A, il calcolo della curva α richiede l'ordinamento dei mesi per intensità crescente delle frequenze, per il ch  si rende necessaria la preliminare scissione delle classi annuali in classi mensili. Questa scissione   stata da noi compiuta molto grossolanamente in considerazione dello scopo esemplificativo del calcolo, ma   evidente che essa   del tutto arbitraria. Tuttavia ogni elaborazione delle serie originali fornite dalle statistiche (nati fra 9-10, 10-11, 11-12 mesi; fra 1-2, 2-3, 3-4, 5-9 anni) pu  evitarsi calcolando anzich  la curva α , la curva di concentrazione δ , la quale ultima non   altro che la curva di graduazione relativa degli eventi secondo l'istante in cui si verificarono. Otte-

(1) C. GINI, *Memorie di Metodologia Statistica*, cit., pag. 387.

TABELLA A. — *Primogeniti nati in Sardegna nel 1936*

DIST. EFFETTIVA		CURVE DI CONCENTRAZIONE						CURVA DI GRADUAZIONE			
Mesi x fra ricorrenza mensile dal ma- trimonio e nascita	Nati nel mese x	Mesi x	Nati negli ultimi x mesi	Curva α		Curva δ		Mesi x fra nona ricorrenza mens. dal matrim. e nascita	Nati al disotto di x mesi	Ordin. % t	Ascisse % n
				Ascisse	Ordin.	Ascisse	Ordin.				
				% p	% q	100- q	100- p			= (8)	= (7)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	905	1	4					1	905		
2	603	2	8					2	1508	3,9	38,1
3	430	3	12					3	1938		
4	133	4	16					4	2071	7,8	52,4
5	133	5	20	9,8	0,5	99,5	90,2	5	2204		
6	133	6	24					6	2337	11,8	59,1
7	133	7	28					7	2470		
8	133	8	32					8	2603		
9	133	9	36					9	2736		
10	133	10	40	19,6	1,0	99,0	80,4	10	2869		
11	133	11	44					11	3002	21,6	75,9
12	133	12	48					12	3135		
13	133	13	55					13	3268		
14	133	14	62					14	3401		
15	133	15	69	29,4	1,7	98,3	70,6	15	3534		
16	24	16	76					16	3558	31,4	90,0
17	24	17	83					17	3582		
18	24	18	90					18	3606		
19	24	19	97					19	3630		
20	24	20	104	39,2	2,6	97,4	60,8	20	3654		
21	24	21	111					21	3678	41,2	93,0
22	24	22	118					22	3702		
23	24	23	125					23	3726		
24	24	24	132					24	3750		
25	24	25	156	49,0	3,9	96,1	51,0	25	3774		
26	24	26	180					26	3798	51,0	96,1
27	24	27	204					27	3822		
28	7	28	228					28	3829		
29	7	29	252					29	3836		
30	7	30	276	58,8	7,0	93,0	41,2	30	3843		
31	7	31	300					31	3850	60,8	97,4
32	7	32	324					32	3857		
33	7	33	348					33	3864		
34	7	34	372					34	3871		
35	7	35	396	68,6	10,0	90,0	31,4	35	3878		
36	7	36	420					36	3885	70,6	98,3
37	7	37	553					37	3892		
38	7	38	686					38	3899		
39	7	39	819					39	3906		
40	4	40	952	78,4	24,1	75,9	21,6	40	3910		
41	4	41	1085					41	3914	80,4	99,0
42	4	42	1218					42	3918		
43	4	43	1351					43	3922		
44	4	44	1484					44	3926		
45	4	45	1617	88,2	40,9	59,1	11,8	45	3930		
46	4	46	1750					46	3934	90,2	99,5
47	4	47	1883	92,2	47,6	52,4	7,8	47	3938		
48	4	48	2016					48	3942		
49	4	49	2446	96,1	61,9	38,1	3,9	49	3946		
50	4	50	3049					50	3950		
51	4	51	3954	100,0	100,0	0,0	0,0	51	3954	100,0	100,0

nuta così la curva di concentrazione δ , si può misurare la concentrazione delle frequenze nel tempo, facendo al solito il rapporto fra l'area di concentrazione e l'area del triangolo ABC (fig. 3).

Per misurare adunque la concentrazione delle nascite dei primogeniti nel tempo, si può, in primo luogo, costruire la curva di graduazione relativa di tali nascite secondo l'istante in cui si verificarono (= curva di concentrazione δ), ponendo ⁽¹⁾:

$$\begin{array}{l} \text{sull'asse} \\ \text{delle ascisse} \\ \text{i valori} \end{array} \left\{ \begin{array}{l} n_i = \frac{N_i}{N} \text{ ossia, percentuale delle nascite verificatesi fino} \\ \text{alla fine dell'i-esimo mese } T_i; (N = N_1 + N_2 + \dots N_u) \end{array} \right.$$

$$\begin{array}{l} \text{sull'asse} \\ \text{delle ordi-} \\ \text{nate i valori} \end{array} \left\{ \begin{array}{l} t_i = \frac{i}{T} \text{ ossia, percentuale del tempo trascorso dalla } 9^a \\ \text{ricorrenza mensile dal matrimonio alla fine dell'i-esimo} \\ \text{mese } T_i \text{ rispetto all'intervallo massimo considerato } T. \end{array} \right.$$

Si calcola poi il valore di R^* in base alla [7], nella quale, al posto di q va sostituito t e al posto di p va sostituito n . La [7] diviene pertanto :

$$R^* = 1 - \sum_{i=0}^n (t_i + t_{i+1}) (n_{i+1} - n_i) \quad [7 \text{ bis}]$$

È questa la formola da noi adoperata nel testo per alcune delle nostre distribuzioni, e della quale è dato un esempio di applicazione concreta alla nota 3 a pag. 16.

8. — IL RAPPORTO DI CONCENTRAZIONE R^* È UN COMBINAZIONE LINEARE DI $I\phi$.

Dalla circostanza che la curva δ (fig. 3) di concentrazione delle frequenze di un evento nel tempo, coincide, per le serie storiche di frequenza *monotone decrescenti*, con la curva di graduazione relativa degli eventi secondo l'istante in cui si verificano, deriva che il rapporto di concentrazione R^* calcolato in base alla [7] o alla [7 bis] per le distribuzioni dei primogeniti, non è altro che una combinazione lineare della media degli intervalli $I\phi$ e dell'ampiezza T del campo di variazione del carattere. Tale ampiezza dev'essere ovviamente fissata in maniera uniforme per tutte le

(1) Cfr. nota 1 a pag. seguente.

distribuzioni da paragonare (nel nostro caso : III mesi, quanti ne corrono fra la 9^a e la 120^a ricorrenza mensile dal matrimonio), e da ciò consegue che, in pratica, R^* è soltanto funzione di $I\phi$.

Si può infatti facilmente dimostrare ⁽¹⁾ che il valore di R^* nella [7 bis] può calcolarsi anche con la formola seguente :

$$R^* = 1 - 2 \frac{M}{T} \quad [8]$$

dove M è la media aritmetica del carattere (distanza media, espressa in mesi, da 9 mesi dopo il matrimonio alla nascita del primogenito, ossia $I\phi - 9$), e T è la ampiezza del campo di variazione del carattere, ampiezza che è stata posta, per le nostre distribuzioni = III mesi.

Si può dunque concludere che per le serie storiche di frequenza *monotone decrescenti* come quelle riguardanti i primogeniti,

(1) Indichiamo con :

N_i = n° assoluto delle nascite verificatesi fino alla fine dell' i -esimo mese T_i

n_i = n° percentuale delle nascite verificatesi fino alla fine dello stesso mese T_i , cioè :

$$n_i = \frac{N_i}{N} \quad (N = N_1 + N_2 + \dots + N_u).$$

t_i = percentuale del tempo trascorso dalla 9^a ricorrenza mensile dal matrimonio alla fine dell' i -esimo mese T_i , rispetto all'intervallo massimo considerato, cioè :

$$t_i = \frac{i}{T}.$$

Il numero concreto dei nati nell' i -esimo mese è $N_i - N_{i-1}$; il numero percentuale è $n_i - n_{i-1}$. Il calcolo della media $M = I\phi - 9$, avviene con la formola :

$$M = \frac{N_1 \cdot 1 + (N_2 - N_1) 2 + (N_3 - N_2) 3 + \dots}{N}.$$

Con le notazioni poste, risulta anche :

$$\begin{aligned} M &= \frac{n_1 N \cdot t_1 T + (n_2 N - n_1 N) t_2 T + (n_3 N - n_2 N) t_3 T + \dots}{N} = \\ &= [n_1 t_1 + (n_2 - n_1) t_2 + (n_3 - n_2) t_3 + \dots] T \end{aligned}$$

da cui risulta :

$$\frac{M}{T} = n_1 t_1 + (n_2 - n_1) t_2 + (n_3 - n_2) t_3 + \dots = \rho$$

il Rapporto di concentrazione nel tempo risulta una combinazione lineare praticamente solo della media $I\bar{p}$ degli intervalli (essendo costante il valore di T) e pertanto il rapporto stesso, non ci può dire molto di più sulla fecondità delle primipare, di quanto già ci dice lo stesso intervallo protogenesico. In altre parole, il valore di R^* per le distribuzioni dei primogeniti sta ad indicare il grado di addensamento delle frequenze verso i primi mesi successivi alla 9^a ricorrenza mensile dal matrimonio: ma tale grado di addensamento è caratterizzato compiutamente dalla media $I\bar{p}$ che rappresenta il baricentro del sistema e della quale R^* non è altro che una combinazione lineare.

9. — LA CONCENTRAZIONE DI UN CARATTERE TEMPORALE PER DISTRIBUZIONI MONOTONE DECRESCENTI, MISURATA CON L'INDICE α DI PARETO.

L'andamento delle curve di distribuzione dei primogeniti in funzione della distanza dal matrimonio, che si può desumere dal Grafico *A*, è di tipo iperbolico e risulta simile all'andamento delle curve di distribuzione dei redditi. Per avere perciò una misura approssimata della rapidità con la quale le nascite diminuiscono a mano a mano che ci si allontana dal matrimonio, si può fare ricorso, alla formola che il Pareto ⁽¹⁾, com'è noto, propose di applicare per descrivere la curva dei redditi globali:

$$y = \frac{K}{x^\alpha} \quad [9]$$

Ma, per la sua stessa definizione:

$$R^* = \frac{\frac{1}{2} - \rho}{\frac{1}{2}} = 1 - 2\rho$$

nella quale, l'espressione ρ rappresenta, con l'approssimazione voluta, l'area racchiusa fra la curva di concentrazione e i due lati del triangolo AFE e BF (fig. 2), mentre $1/2$ rappresenta l'area del triangolo ABF . Sicchè il numeratore della frazione è l'area racchiusa fra la curva e la retta di equidistribuzione (area di concentrazione), mentre tutta la frazione rappresenta il rapporto fra quest'area e quella del triangolo ABF .

Sostituendo a ρ il suo valore, si ha:

$$R^* = 1 - 2 \frac{M}{T}.$$

(1) V. PARETO, *Cours d'Economie Politique professé à l'Université de Lausanne*, 1897.

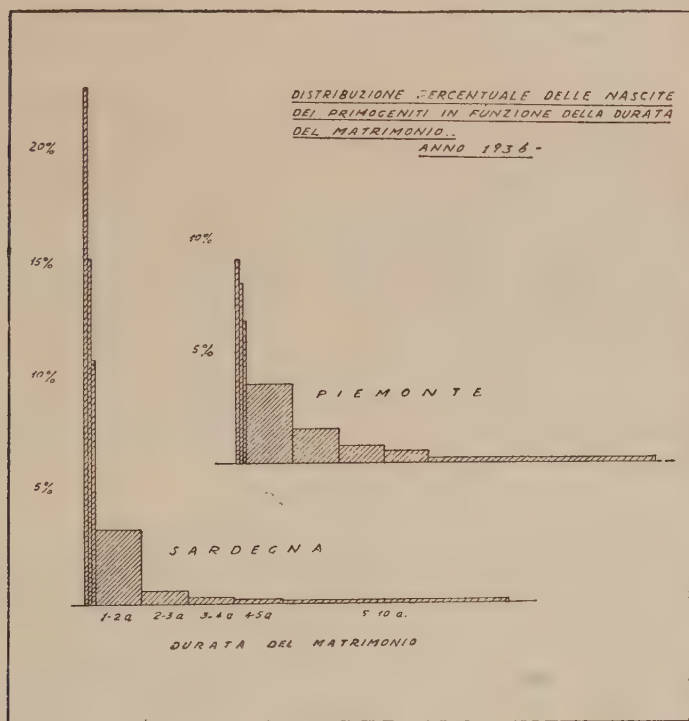
nella quale, y è il numero degli individui che posseggono un reddito superiore a x (limite inferiore di ogni classe di reddito), K è una costante da determinarsi di volta in volta, e α è l'indice di distribuzione.

Passando ai logaritmi si ha la relazione:

$$\log y = \log K - \alpha \log x \quad [10]$$

la quale, in un diagramma a doppia scala logaritmica, è rappresentata da una retta avente sull'asse x l'inclinazione α , e può con-

GRAFICO A



venientemente rappresentare la spezzata rettilinea, molto simile a una retta, che si ottiene se, anche su un diagramma a doppia scala logaritmica, si segnano sulle ascisse i logaritmi delle classi e sulle ordinate i corrispondenti logaritmi del numero degli individui che posseggono al disopra del limite inferiore di ogni classe. Ciò, però, si verifica solo per le seriazioni di redditi globali, mentre

per altre seriazioni (patrimoni, successioni, redditi da patrimonio ecc.), le spezzate risultano alquanto concave rispetto all'asse x .

Per i redditi, l'inclinazione α indica la rapidità con la quale diminuiscono i redditieri, di mano in mano che si passa dai bassi redditi a quelli più elevati: una inclinazione forte indicherebbe che i redditieri, nelle classi via via più alte, vanno rapidamente diminuendo. Al contrario, l'abbassarsi lento della retta, indica che nelle alte classi di reddito, la diminuzione del numero di coloro che li posseggono, avviene lentamente. Un alto valore di α indica debole concentrazione dei redditi; e viceversa, basso valore di α , indica forte concentrazione dei redditi stessi ⁽¹⁾.

Nel caso delle nostre distribuzioni, ponendo:

y = numero dei primogeniti nati a distanza superiore ad x mesi dal matrimonio;

x = distanza in mesi tra il matrimonio e la nascita (limite inferiore di ogni classe di durata di matrimonio),

abbiamo potuto verificare (cfr. più innanzi, n. 11) che le spezzate logaritmiche risultanti, pur presentando una certa convessità o concavità verso l'asse delle x , si accostano abbastanza bene ad una retta, e possono quindi, con una certa approssimazione, essere rappresentate appunto da una retta inclinata sull'asse delle x .

10. — LA CONCENTRAZIONE DI UN CARATTERE TEMPORALE PER SERIE DI FREQUENZA MONOTONE DECRESCENTI, MISURATA CON GLI INDICI R E δ DEL GINI.

Le distribuzioni dei primogeniti in funzione della distanza dal matrimonio possono essere anche studiate applicando con immediatezza formale i procedimenti che danno la misura della concentrazione o variabilità relativa dei caratteri.

In tal caso, riferendosi alla fig. 3, sull'asse delle ascisse sarà indicato il numero (relativo) dei primogeniti nati entro un certo

(1) Sul significato e sulla interpretazione delle variazioni di α (quando α aumenta, diminuisce la concentrazione e viceversa) cfr.: C. GINI, *Memorie di Metodologia Statistica*, cit., pag. 33 e segg.; A. NICEFORO, *Il Metodo Statistico*, Nuova Edizione, pag. 474 e seg. G. DE MEO, *Distribuzione della ricchezza e composizione demografica in alcune città dell'Italia Meridionale alla metà del Secolo XVIII*, in: « Annali di Statistica », Serie VI, Vol. XIX, 1931, pag. 34-36.

intervallo dal matrimonio, mentre sull'asse delle ordinate verrà rappresentato l'*ammontare del carattere* che compete a quella frazione di primogeniti, (somma degli intervalli fra 9 mesi dopo il matrimonio e la nascita del primogenito). Così operando, si ottiene la curva di concentrazione degli intervalli da 9 mesi dopo il matrimonio alla nascita del primogenito che, quindi, ha un significato affatto diverso — come ora meglio vedremo (cfr. n. 11) — dalla curva di concentrazione di cui al n. 7.

Per il calcolo del rapporto di concentrazione, si può usare la formola del Gini ⁽¹⁾

$$R = \frac{\sum_{i=1}^s (i_i + i_{i-1} - 1) S_i}{(n - 1) A_n} - 1 \quad [11]$$

nella quale, i_i indica il numero degli individui nati al di sotto del tempo x_i (dalla 9ª ricorrenza mensile dal matrimonio); S_i , l'ammontare del carattere (prodotto: tempo \times nascite) cadente nella classe ell-esima; n , il totale dei primogeniti, ed A_n la somma degli S_i , ossia l'ammontare totale del carattere (somma dei prodotti: tempo \times nascite per tutti i primogeniti considerati).

Potrebbe usarsi anche la [7 bis], nella quale, a t_i e t_{i+1} , deve sostituirsi rispettivamente Q_i e Q_{i+1} ; Q_i essendo, con le notazioni della nota (1) a pag. 149, $= \frac{i N_i}{N}$. Per cui la [7 bis] diviene:

$$R = 1 - \sum_{i=0}^N (Q_i + Q_{i+1}) (n_{i+1} - n_i) \quad [12]$$

Abbiamo verificato numericamente la identità di risultati cui si perviene con le formole [11] e [12]. Applicando la formola [12] basata su valori percentuali e non su valori assoluti, si ottiene una minore approssimazione, ma si consegue una certa economia di tempo nel calcolo dell'indice.

I valori di R^* per le distribuzioni dei primogeniti calcolati con la [7 bis] dànno, come abbiamo visto, una misura della variabilità relativa del carattere: *frequenza delle nascite nelle diverse unità temporali* (i vari mesi che seguono il matrimonio). Pertanto, in questo caso, un elevato valore di R^* sta a testimoniare che le

(1) C. GINI, *Memorie di Metodologia Statistica*, cit., formola 13 a pag. 366.

nascite si addensano fortemente sui primi mesi che seguono la 9ª ricorrenza dal matrimonio stesso: vale a dire che R^* , in generale, tende ad aumentare quando le frequenze delle nascite, tendono ad addensarsi su un'unica unità temporale (la prima, nella successione naturale delle unità).

Con le formole [11] e [12], invece, si perviene ad una misura della variabilità relativa del carattere: *intervallo tra matrimonio e nascita*. In questo secondo caso, il valore di R tende ad aumentare quando aumenta il numero relativo delle nascite che avvengono a grande distanza dal matrimonio: così come, per i redditi, il rapporto di concentrazione tende in generale ad aumentare quando aumenta la percentuale dei redditieri con reddito più elevato.

Volendo misurare la concentrazione di un carattere temporale, quale l'istante in cui si verifica un dato evento misurato a partire da una certa origine, si può fare ricorso anche alla [3] che fornisce il valore dell'indice di concentrazione δ del Gini, oppure all'indice α di Pareto (cfr. n. 9).

II. — ESEMPI DI APPLICAZIONE DEI PRECEDENTI INDICI.

Allo scopo di metter meglio in evidenza il significato dei vari indici che possono usarsi per la misura della concentrazione nel tempo delle nascite dei primogeniti, abbiamo costruito la seguente Tabella B nella quale, per l'anno 1936, riportiamo, Compartimento

TABELLA B

Indici di distribuzione e di concentrazione delle nascite dei primogeniti nel tempo.

	I_p	R^* formola [7 bis]	α formola [10]	R formola [11] o [12]	δ formola [3]
Piemonte	26,82	67,1	— 1.254	35,6	1.859
Liguria	23,33	73,6	— 1.460	34,6	1.842
Lombardia	22,51	75,2	— 1.526	33,8	1.832
Venezia Tridentina . . .	19,11	81,5	— 1.803	31,4	1.767
Veneto	17,81	83,9	— 1.997	28,9	1.684
Venezia Giulia	20,74	78,4	— 1.640	33,1	1.818
Emilia	20,72	78,5	— 1.731	31,3	1.734
Toscana	19,82	79,3	— 1.733	30,8	1.756
Marche	17,90	83,8	— 2.005	28,7	1.655
Umbria	19,60	80,6	— 1.871	29,7	1.701
Lazio	18,09	83,0	— 1.908	29,9	1.724
Abruzzi	19,25	81,2	— 1.866	30,2	1.701
Campania	18,18	83,2	— 1.979	29,1	1.698
Puglie	18,44	82,8	— 1.912	30,1	1.697
Lucania	18,02	84,0	— 2.130	28,3	1.599
Calabria	18,13	83,3	— 1.978	29,2	1.676
Sicilia	19,61	80,5	— 1.821	30,5	1.732
Sardegna	17,30	84,8	— 2.086	27,8	1.675

per Compartimento, i valori: $I\phi$ (intervallo protogenesico), R^* (formola [7 bis]), α (formola [10]), R (formola [11] o [12]), e δ (formola [3]). Per il calcolo di tutti questi valori sono state considerate soltanto le nascite dei primogeniti avvenute dopo la 9ª ricorrenza mensile, entro i primi 10 anni dal matrimonio ⁽¹⁾.

Le reciproche relazioni fra questi vari indici sono abbastanza bene messe in evidenza dagli indici di cograduazione riportati nella seguente Tabella C:

TABELLA C

Indici di cograduazione I fra gli indici di cui alla Tabella B.

Cograduazione fra	I
$I\phi$ e R^*	— 0,94
$I\phi$ e R	+ 0,83
$I\phi$ e δ	+ 0,78
$I\phi$ e α	— 0,86
R^* e R	— 0,85

Il valore di $I = -0.94$ fra $I\phi$ ed R^* ci dice che, a mano a mano che l'intervallo medio si fa più piccolo, aumenta la concentrazione nel tempo, ossia le nascite dei primogeniti vanno addensandosi di preferenza sulle brevi distanze dal matrimonio. La cograduazione, anzi, dovrebbe in questo caso, risultare perfetta ($I = -1$), in quanto R^* , per ciò che abbiamo detto, è una combinazione lineare di $I\phi$: se, nel caso concreto, ciò non si verifica, è a causa delle approssimazioni intervenute nella determinazione delle due serie di indici.

Il valore $I = +0.83$ fra $I\phi$ ed R ci dice al contrario che, aumentando l'intervallo medio, aumenta la variabilità della massa, in quanto, in questo caso, sono più largamente rappresentati gli intervalli lunghi, la cui presenza, influisce appunto nel senso di accentuare la variabilità relativa della massa studiata, così come accade per i redditi, la cui concentrazione aumenta, in generale, quando sono relativamente più numerosi i redditi più elevati. Analoga interpretazione suggerisce il valore di $I = +0.78$ fra

(1) È questa la ragione per cui i valori di $I\phi$ riportati nella Tabella B non coincidono con quelli riportati nel testo alla Tabella II, i quali ultimi furono ottenuti in base a tutte le nascite avvenute dopo la 9ª ricorrenza mensile, comprese quindi quelle verificatesi dopo i dieci anni dal matrimonio.

$I\phi$ e δ , δ essendo una quantità che cresce con l'aumentare della concentrazione, ossia, della variabilità relativa.

Fra $I\phi$ ed α esiste forte cograduazione negativa ($I = -0,86$). Nel caso delle distribuzioni dei primogeniti, un elevato valore di α sta infatti ad indicare che i primogeniti, a mano a mano che ci si allontana dalla data del matrimonio, vanno rapidamente diminuendo. L'abbassarsi lento della retta — piccolo valore di α — sta ad indicare che i primogeniti, a mano a mano che ci si allontana dalla data del matrimonio, diminuiscono lentamente. Nel primo caso (alto valore di α) si ha dunque forte addensamento delle nascite alle brevi distanze dal matrimonio (il che implica piccolo valore di $I\phi$); nel secondo (basso valore di α) si ha sensibile diluizione delle nascite stesse lungo tutto il periodo considerato (il che implica più elevato valore di $I\phi$).

Fra R^* ed R , infine, esiste forte cograduazione negativa ($I = -0,85$). Ciò avviene perchè, in generale, quando si accentua lo addensamento delle frequenze sulle brevi distanze dal matrimonio; diminuisce la variabilità relativa della massa e viceversa.

12. — RIASSUNTO E CONCLUSIONI.

Riassumendo le considerazioni fatte nella presente Appendice, possiamo dire che la concentrazione di un evento nel tempo, ed in particolare la concentrazione nel tempo di eventi relativi a serie storiche di frequenza non mai crescenti, può essere studiata con differenti criteri.

Se si vuole avere un'idea del grado di addensamento delle frequenze su alcune più che su altre unità temporali, si può fare ricorso all'indice R^* che è il rapporto di concentrazione del carattere: *frequenza degli eventi nelle varie unità temporali considerate*. Tale indice di concentrazione, nel caso particolare di serie storiche di frequenza non mai crescenti, risulta una combinazione lineare della media aritmetica degli intervalli di tempo tra un'origine comune e i vari istanti nei quali i singoli eventi si producono. Pertanto, l'indice R^* calcolato per le distribuzioni dei primogeniti, è una combinazione lineare dell'intervallo protogenesico $I\phi$.

Se si adotta per le stesse serie storiche di frequenza il criterio usuale di concentrazione, l'indice R è il rapporto di concentrazione del carattere: *intervallo di tempo fra un'origine comune e i vari istanti nei quali gli eventi si producono*. Per le distribuzioni dei

primogeniti, R fornisce dunque la misura della variabilità relativa degli intervalli fra 9^a ricorrenza mensile dal matrimonio e la nascita del primogenito. Qualche cosa di analogo avviene se, in luogo di R del Gini si fa uso dell'indice δ dello stesso A .

Altro criterio, infine, per giudicare del grado di addensamento delle nascite dei primogeniti sui brevi intervalli dal matrimonio, consiste nel calcolare per le distribuzioni stesse, l'indice di distribuzione α del Pareto, che, col suo valore, sta ad indicare appunto la rapidità con la quale diminuisce il numero dei primogeniti a mano a mano che ci si allontana dalla data del matrimonio.

C. EVELPIDI

Le revenu national des pays balkaniques

Introduction. — Le revenu national a fait l'objet d'un certain nombre d'évaluations dans les pays balkaniques. Mais les résultats obtenus par les divers auteurs présentent des écarts appréciables à cause, surtout, du manque de données fiscales pouvant servir de base à ces évaluations. En effet, le plus clair du revenu national est fourni dans ces pays par des millions de petites exploitations agricoles, appartenant aux paysans, après surtout la réforme agraire introduite dans ces pays depuis la guerre. Ces exploitations ne sont pas soumises à l'impôt sur le revenu et une quantité importante de leur production est consommée par les agriculteurs mêmes, sans réaliser un revenu en espèces, susceptible d'une estimation fiscale, commerciale ou autre. Nous rencontrons des difficultés analogues dans l'estimation de la production artisanale qui, dans ces pays, a parfois une importance égale à celle de l'industrie, laquelle s'y trouve en voie de développement. L'impôt sur le revenu ne peut donc servir de base pour l'évaluation ou tout au moins pour le contrôle du revenu national. Aussi bien tous ceux qui se sont occupés de l'évaluation de ce revenu dans les Etats balkaniques se sont vus obligés de suivre la méthode objective. Toutefois, les difficultés que présente cette méthode se trouvent renforcées notamment par les lacunes fréquentes des statistiques servant de base à une telle évaluation. L'interprétation des données existantes ou leur rectification donne des résultats différents, suivant les auteurs qui se sont occupés du calcul du revenu national dans les pays précités. D'où la nécessité de reprendre les calculs pour une série d'années, là

où cela est possible. D'autant plus qu'on ne saurait se baser sans risque sur le résultat d'une seule année, la production agricole de ces pays étant sujette à des fluctuations annuelles importantes et les prix de cette production ayant oscillé énormément, durant surtout les années de la crise agricole.

Le revenu national de la Grèce. — Les estimations du revenu national faites en Grèce ont donné les résultats suivants, en millions de drachmes :

X. Zolotas	46.913 millions pour l'année 1927 ⁽¹⁾
P. Rediadis	34.700 » » » 1929 ⁽²⁾
P. Dertilis	45.000 » » » 1928 ⁽³⁾

Il y a lieu de remarquer que la drachme a été stable durant les années considérées et que les effets de la crise mondiale n'avaient pas encore paru en Grèce. Il n'est pas sans intérêt de noter enfin que M. C. Athanassiadis a calculé le revenu national de la Grèce à 51.341 millions, en se basant sur l'équation de Pareto ⁽⁴⁾ ; mais il faut prendre en considération que l'application de l'équation augmente, selon M. J. Stamp, le résultat obtenu ⁽⁵⁾.

Nous avons repris le calcul, après contrôle et rectification des données statistiques et en ajoutant au revenu national certains postes omis dans les évaluations précédentes (industrie paysanne, métiers, bâtiment, manipulation de tabacs, revenus non taxés des forêts) ⁽⁶⁾. D'après nos calculs, le revenu national de la Grèce a oscillé comme il suit, en millions de drachmes, durant la période décennale 1927-36 :

(1) X. ZOLOTAS, *Stabilisation monétaire* page 98, Athènes, 1929 (en grec).

(2) P. REDIADIS, *The Greek national Income and Wealth in 1929*, « Metron », Vol. VIII, N. 4, Roma, 1930.

(3) P. DERTILIS, *Le revenu nat. de la Grèce* dans « Archive des Sciences économiques et sociales ». Vol. 12, page 351, Athènes 1932 (en grec).

(4) C. ATHANASSIADIS, *La courbe des revenus* dans « Revue des Sciences économiques et financières ». Vol. I, N. 3, Athènes, 1932 (en grec).

(5) J. STAMP, *Methods used in different countries for estimating National Income*. « Journal of the Royal Statistical Society ». Vol. XCVII, pages 429, 1934.

(6) C. EVELPIDI, *Le revenu national*. Athènes, 1937 (en grec).

1) *Production naturelle.*

	Année 1927	Année 1936	Moyenne 1927-36
Agriculture	9.522	15.235	11.036
Elevage	4.185	4.619	3.903
Forêts	999	1.200	1.071
Mines et salines	306	466	331
Pêche et chasse	450	480	445
	15.462	22.000	16.786

2) *Transformation.*

Industrie paysanne	550	640	530
Métiers et bâtiment	3.000	3.350	2.995
Industrie	3.875	6.190	4.440
	7.428	10.180	7.965

3) *Circulation.*

Transports terrestres	1.946	2.630	2.254
Transports maritimes	588	1.467	811
Postes, télégraphes, téléphones	175	333	226
Commerce	4.946	6.263	4.933
	7.655	10.693	8.224

4) *Gains de capitaux.*

Banques et Assurances	1.817	2.645	2.090
Loyers	3.870	4.200	3.870
Titres de l'Etat	900	972	1.026
Envois des émigrés	2.595	1.938	2.350
	9.186	9.755	9.336

5) *Services.*

Etat et communes	2.285	3.444	2.777
Services personnels	500	870	668
Professions libres	1.550	2.426	1.878
	4.335	6.740	5.323
Totaux généraux	44.062	59.268	47.634

L'apport relatif des branches principales se répartit donc comme il suit :

Production naturelle	35 -	37,2	35,2
Transformation	16,9	16,9	16,8
Circulation	17,4	18,1	17,3
Gains de capitaux	20,9	16,4	19,5
Services	9,8	11,4	11,2
	100 -	100 -	100

Ce revenu correspond en moyenne à 8.604 drachmes par tête d'habitant en 1936, contre 7.107 en 1927. Il présente donc une augmentation de 21,8 %, tandis que pendant ce même laps de temps l'augmentation du coût de la vie a été de 13 % et celle des prix de gros de 17,91 %.

Bulgarie. — Le revenu national de la Bulgarie a été calculé par M. le Dr. Assen Tchakaloff pour une série d'années allant de 1924 à 1935 ⁽¹⁾. La répartition de ce revenu est la suivante, en millions de levas :

	1927	1928	1929	1934	1935
Economie rurale.	30.826	33.637	32.247	17.578	19.565
Artisanat.	4.309	4.537	4.704	2.977	2.816
Industrie.	2.751	2.920	3.017	2.482	2.272
Communications et transports	1.076	1.173	1.213	1.088	1.045
Commerce.	3.791	4.024	4.263	2.181	2.481
Crédit et assurances.	1.496	1.769	2.058	1.344	1.355
Prof. libérales.	570	580	600	320	302
Immeubles.	2.660	2.845	2.988	2.353	2.177
Traitements.	4.689	4.779	4.794	3.987	4.174
Entreprises publiques.	294	265	323	254	382
Totaux.	52.462	56.529	56.207	34.564	36.569

En pourcentage, la répartition du revenu bulgare a été comme il suit :

	1927	1928	1929	1934	1935
Economie rurale.	58,8	59,5	57,4	50,9	53,5
Artisanat et industrie.	13,4	13,1	13,8	15,8	13,2
Professions libérales et traitements.	10 —	9,5	9,6	12,4	12,2
Autres revenus.	17,8	17,9	19,2	20,9	21,1
	100	100	100	100	100

Ce revenu nous donne une moyenne par tête d'habitant de 9512 levas en 1927, 10.115 levas en 1928, 5965 levas en 1935. Calculé d'après les indices des prix achat, ce revenu se monte :

en 1927	à	9.512
1928	»	9.673
1929	»	9.498
1934	»	8.280
1935	»	8.837

(1) AS. TCHAKALOFF, *The national Income of Bulgaria 1924-1935* ; SUMMARY, *Publications of the Statistical Institute for economic research. State University of Sofia. Sofia, 1937.*

Roumanie. — Le calcul du revenu national de ce pays a été fait comme il suit, en millions de lei, pour les années 1928-33 ⁽¹⁾ :

1) *Biens*

	1928	1929	1930	1931	1932
Céréales et autres plantes. . .	81,5	78,8	40,3	37 -	37,5
Elevage	41,3	39,7	33 -	24 -	21 -
Forêts	7 -	6 -	5,5	4,5	3,5
Economie rurale. . .	129,8	124,5	78,8	65,5	62 -
Industrie et mines.	33,3	36 -	34,1	19,6	18 -
Constructions	5,7	2,8	2,5	2 -	3 -
	39 -	38,8	36,6	21,6	21 -
Total des biens . . .	168,8	163,3	115,4	87,1	83 -

2) *Services*

	1928	1929	1930	1931	1932
Loyers	19,1	19,3	18 -	15,1	14 -
Transports et communications	2,3	2,2	2 -	1,8	1,2
Banques, commerce, assurances	4,1	4,5	4 -	1 -	1 -
Services professionnels . . .	5,6	5,6	4,5	4,8	3,7
Services domestiques. . . .	1 -	1 -	1 -	0,8	0,6
Total des services . . .	32,1	32,6	29,5	23,5	20,5
Total général	200,9	195,9	144,9	110,6	103,5

Pour l'année 1933 le revenu national a été évalué globalement à 99,3 milliards de lei.

L'apport relatif des différentes branches du revenu national a été évalué comme il suit :

	1928	1932
Economie rurale.	64,5 %	59,9 %
Industrie, mines, constructions	19,5 %	20,3 %
Services professionnels et domestiques	3,3 %	4,1 %
Autres services	12,7 %	15,7 %
	100	100

L'indice du revenu national a donc baissé de 100 en 1928 à 51,7 en 1932 et 49,4 en 1933. Tandis que l'indice des prix de gros a évolué de 100 en 1929 à 54 en 1932 et 52,3 en 1933 et celui

(1) *Le revenu national de la Roumanie*, dans « Bulletin d'information et de documentation de la Banque Nationale de Roumanie ». Bucarest, Juillet 1930. — V. SLAVESCO, *La situation économique de la Roumanie et sa capacité de paiement*. Bucarest 1934, pages 50-53.

des prix de détail de 100 en 1929 à 63 et à 58 pour ces mêmes années. Rapporté à la population, ce revenu représente par tête d'habitant 11.546 lés en 1928, 5610 en 1932 et 5.327 en 1933, qui représentent respectivement 8765 et 8980 unités, comparative-ment à la valeur d'achat du léi en 1929.

Turquie. — Le revenu national de la Turquie a été calculé par le Service de conjoncture de la République Turque, qui a publié une étude analytique à ce sujet concernant les années 1933-34, 1934-35 et 1935-36 ⁽¹⁾. D'après cette étude, ce revenu serait par branches principales le suivant, en livres turques :

	1933-34	1934-35	1935-36
Agriculture	220,1	263,3	296,6
Élevage	113 -	107,9	108,4
Forêts	23,8	24,9	26,8
Chasse et pêche	2,9	3,1	3,2
Industries et métiers	180 -	200 -	215 -
Loyers	60 -	61,1	62,2
Services publics et privés.	100 -	101,8	103,6
Commerce, communications, assu- rances, banques, etc.	449,9	488 -	514,5
	1149,7	1250,1	1330,3

En pourcentage, ce revenu serait comme il suit :

	1933-34	1934-35	1935-36
Agriculture et élevage	28,9	29,7	30,3
Forêts, chasse et pêche.	2,3	2,2	2,3
Industries et métiers	15,7	16,0	16,2
Loyers	5,2	4,9	4,7
Services	8,7	8,1	7,8
Autres revenus	39,2	39,1	38,7
	100	100	100

Ce revenu donne en moyenne par tête d'habitant 73,56 livres turques en 1933-34, 78,54 en 1934-35 et 82,10 en 1935-36 ;

(1) T. C. *Ekonomi Bakanligi, Konjonktur Servisi*, Seri : -1 No. 7 : *Türkiye Milli Geliri*, Ankara, 1937. Un résumé de cette étude, écrite en langue turque, a été publié en allemand par M. F. EPPENSTEIN : *Das türkische Volkseinkommen*, Prague 1938.

mais calculé en livres de la valeur d'achat de 1933-34 il revient pour ces mêmes années à 73.56, 75.37 et 75.61.

Yougoslavie. — Toutes les études sur le revenu national parues durant les dix dernières années se sont basées sur le travail minutieux qu' a publié en 1927 un groupe de cinq économistes distingués ⁽¹⁾. D'après ce travail, le revenu national de ce pays serait en 1926 de 69.608 millions de dinars. En 1931, ce revenu a été calculé à 42.041 dinars. Enfin une commission de spécialistes, nommée au Ministère des Finances, a évalué à 34.225 millions en 1935 et 38.280 millions en 1936 le revenu national de la Yougoslavie.

D'après les travaux précités, nous pouvons donc dresser le tableau analytique ci-dessous, dans lequel les chiffres expriment des millions de dinars :

	1926	1931	1935	1936
Agriculture	18.638,9	9.191 -	8.325 -	11.359,3
Elevage	12.739,9	5.815,7	6.890,8	7.556,3
Apiculture, sériciculture . .	116,1	52,8	64,4	98,8
Pêche et chasse	167,8	96,5	89,2	98,1
Sylviculture	3.935,6	2.637,0	1.939 -	1.970 -
Mines	900 -	951,1	897,1	898,7
Economie ménagère et domestique	4.837 -	3.627,8	3.386 -	3.100 -
Industrie	5.823,3	3.500 -	3.010 -	3.173,3
Métiers	7.590 -	5.214 -	2.954,7	3.043,5
Commerce	8.480 -	6.511,5	2.676,4	2.748,9
Institutions de crédit . . .	525 -	657,2	323 -	432,5
Sociétés d'assurances . . .	150 -	68 -	35,9	39,8
Chemins de fer	1.796,8	1.396,1	1.394,8	1.319,7
Marine marchande	316,1	316,1	326,1	287,7
Marine fluviale	150 -	120 -	93,3	92,3
Transports en automobiles .	300 -	300 -	231 -	282,1
Autres transports	2.032,5	664,5	730,9	730,9
Postes, télégraphes, téléphones	300,3	293,4	305,1	354,7
Bâtiments	808,9	728 -	582,4	694 -
	69.608,2	42.040,7	34.255,1	38.280,6

(1) *Notre économie nationale et notre revenu national*, par V. M. DJURICIC, M. B. TOSIC, A. VEGNER, P. RUDCENKO et Dr. M. R. DJORDJEVIC, Serajevo, 1927.

Les fluctuations de ce revenu en pourcentage sont donc les suivantes, par branches principales :

	1926	1931	1935	1936
Economie rurale.	48,4	39,4	47,1	51,0
Industrie et métiers	21,3	23,1	21,3	19,5
Autres revenus	30,3	37,5	31,6	29,5
	100	100	100	100

Mais il y a lieu de remarquer que dans le calcul ci-dessus ne sont pas compris les appointements des fonctionnaires et les gains des professions libérales. Nous trouvons les premiers dans les budgets de l'Etat yougoslave et nous évaluons les seconds, d'après le nombre des personnes s'y adonnant et leurs revenus moyens annuels. Nous arrivons ainsi aux chiffres suivants :

	1926	1931	1935	1936
Emoluments des fonctionnaires	6.820	6.240	5.005	5.010
Gains de professions libérales.	2.100	1.500	1.200	1.300
	8.920	7.740	6.205	6.310
Totaux généraux	78.528	49.780,7	40.460,1	44.590,6

Ce revenu correspond à 5917 dinars par tête d'habitant en 1926, à 3581 dinars en 1931, à 2734 dinars en 1935 et à 2973 dinars en 1936. Les indices des prix de gros ont évolué de 100 en 1929, année de base, à 65,2 en 1931, à 55,8 en 1935 et à 62,0 en 1936, tandis que ceux du coût de la vie ont baissé de 100 en 1929 à 83,7 en 1931, à 60,6 en 1935 et à 61,1 en 1936. Le revenu moyen par tête d'habitant se montait donc à 4.313 unités en 1931, à 4.511 unités en 1935 et à 4.874 unités en 1936, comparativement à la valeur d'achat de la monnaie nationale yougoslave en 1929.

Quelques remarques. — Il est un peu risqué de procéder à des comparaisons entre les données ci-dessus, vu les méthodes différentes employées pour l'évaluation du revenu national dans chacun des pays précités, d'autant plus que les années prises en considération ne correspondent pas toujours entre elles. Toutefois nous avons cru utile de procéder à quelques confrontations de ces données à titre tout à fait indicatif et d'en tirer non pas des

conclusions mais quelques remarques, qui présentent un certain intérêt pour l'étude de l'économie des pays balkaniques, comme nous l'avons déjà fait pour le produit national des Etats balkaniques (1).

Le revenu national en francs or a été calculé comme il suit sur la base de la période triennale 1933-35 pour la Grèce, la Turquie et la Bulgarie, pour la période biennale 1935-36, en ce qui concerne la Yougoslavie et pour la période 1928-33, en ce qui concerne la Roumanie.

	Revenu moyen		Par tete d'habitant			
	millions	frs	or	160	frs	or
Turquie	2.998	»	»	»	199	»
Yougoslavie	3.010	»	»	»	»	»
Bulgarie	1.274	»	»	»	210	»
Grèce	1.470	»	»	»	222	»
Roumanie	4.482	»	»	»	247	»

Ce revenu est donc assez bas pour tous les pays balkaniques, restant toujours bien au dessous de celui des pays de l'Europe centrale ou occidentale.

Quant à sa répartition, elle a été calculée comme il suit, par branches principales :

	Production agraire	Industrie et métiers	Autres Revenus
Bulgarie	51,4	15,6	43 -
Grèce	37 -	15,5	47,5
Roumanie	60,7	21,0	18,3
Turquie	31,3	15,7	53 -
Yougoslavie	49,1	20,4	30,5

L'économie nationale de ces pays se base donc surtout sur leur production agricole. C'est pour ces motifs qu'ils ont vu leur revenu national s'abaisser énormément durant la crise agraire (1929-34). Ce revenu remonte depuis, par suite de la hausse des prix des produits agricoles.

Le pourcentage du revenu provenant de l'industrie et des métiers se trouve à peu près au même niveau dans les Etats examinés. Il faut remarquer toutefois que leur industrialisation fait des progrès sensibles d'année en année, ce qui provoque une

(1) C. EVELPIDI, *Les Etats Balkaniques*, page 341, Paris, 1930.

augmentation des revenus industriels de ces pays, aux dépens surtout des revenus de l'artisanat.

Les émoluments des fonctionnaires, des employés et des professions libérales ont présenté, en chiffres absolus, une certaine diminution durant la crise, mais cette diminution n'a pas été proportionnelle à la baisse du revenu national, d'où une augmentation de l'importance en pourcentage de ces postes, chose qui a été d'ailleurs observée dans presque tous les pays ⁽¹⁾. De même les frais des comptes budgétaires ne sont pas assez compressibles pour suivre la diminution du revenu national, d'où une augmentation relative de la part y prélevée par le fisc, les communes et les organisations de droit public ; elle a été en 1935 de 26,2 % pour la Grèce contre 22,9 % en 1929, de 31,51 % en 1935 pour la Bulgarie contre 20,73 % en 1927, de 26,6 % pour la Yougoslavie en 1935 contre 20,9 % en 1931 et de 26,54 % pour la Roumanie en 1933 contre 20,58 % en 1929.

En moyenne annuelle pendant la période 1934-36, elle a été de 17,2 % pour la Turquie, de 26,8 % pour la Yougoslavie (1925-36), de 26,9 % pour la Grèce et de 30,27 % pour la Bulgarie. Exception faite pour ce dernier Etat, cette proportion n'est pas aussi forte que dans les pays riches de l'Europe, mais il y a lieu de considérer qu'elle est prise sur un revenu déjà restreint, qui ne laisse pas de marge à une plus forte imposition. C'est pourquoi ces Etats sont obligés d'avoir recours surtout aux impôts indirects, vu que l'imposition du revenu personnel, deduction faite des minima non taxés, ne pourrait pas leur fournir les fonds nécessaires à leurs dépenses.

Une question d'une importance capitale pour les pays balkaniques est la répartition du revenu entre la population urbaine et la population rurale. Celle-ci constitue la grande majorité des habitants de ces pays, puisqu'elle en comprend les 80 % en Turquie et en Roumanie, les 77 % en Bulgarie, les 67 % en Yougoslavie et en Grèce ; mais le total des revenus de l'économie rurale (agriculture, élevage, forêts, pêche et chasse, industrie paysanne) se monte seulement à 60,7 % en Roumanie (1928-33), 51,4 % en Bulgarie, 49,1 en Yougoslavie, 37,0 en Grèce, 31,3 % en Turquie, pour les périodes des années mentionnées ci-dessus.

(1) WL. WOYTINSKY, *Les conséquences sociales de la crise*, page 275, Genève, 1936.

D'où la grande différence de revenu entre la population urbaine et la population rurale des pays balkaniques, ainsi qu'il appert du tableau ci-dessous.

	Population rurale	Revenus de l'économie rurale	Population urbaine	Revenus l'économie urbaine
	Milliers d'habitants	Millions Frs. Or.	Milliers d'habitants	Millions Frs. Or.
Bulgarie	4.689	655	1.401	619
Grèce	4.520	544	2.226	926
Turquie	12.728	937	3.182	2.061
Roumanie	14.448	2.721	3.612	1.761
Yougoslavie	10.117	1.478	4.983	1.532

Il revient donc par tête d'habitant, en francs or :

	Population rurale	Population urbaine
Bulgarie	140	442
Grèce	120	415
Turquie	74	648
Roumanie	188	489
Yougoslavie	145	306

La crise agricole ayant amené une forte diminution des revenus de l'économie rurale des pays balkaniques, exportateurs de produits agricoles, il s'ensuivit une grande misère de leur population rurale. En revanche les produits de leur industrie n'ont pas présenté une grande baisse des prix, car ils ont pu être protégés efficacement, vu que cette industrie travaille, d'une façon générale, pour les besoins du marché intérieur qu'elle ne peut pas couvrir. D'où les mesures exceptionnelles prises dans ces pays pendant la crise, afin de relever le revenu des agriculteurs, c.à.d. de tenter une répartition plus équitable du revenu national et même, en partie, du capital national, à laquelle visaient les lois sur le règlement des dettes agricoles.

